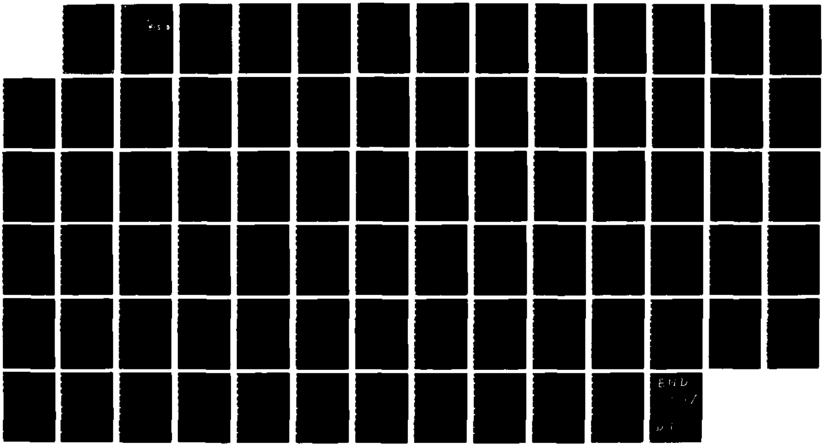
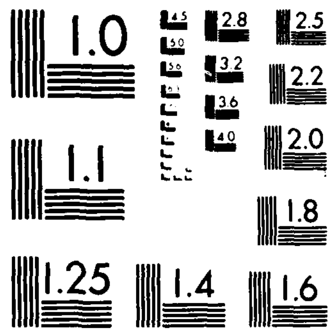
AD-A184 826 A CLUSTER ANALYSIS OF MANPOWER SUPPLY TO ARMY RESERVE CENTERS(U) NAVAL POSTGRADUATE SCHOOL MONTEREY CA A R JONES JUN 87 1/1

UNCLASSIFIED F/G 15/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(2)

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

A CLUSTER ANALYSIS OF MANPOWER SUPPLY TO
ARMY RESERVE CENTERS

by

Alvin R. Jones

June 1987

Thesis Advisors: George W. Thomas
Stephen L. Mehay

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b DECLASSIFICATION / DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION Naval Postgraduate School 6b OFFICE SYMBOL (If applicable) Code 54	7a NAME OF MONITORING ORGANIZATION Naval Postgraduate School
6c ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000	7b ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000
8a NAME OF FUNDING / SPONSORING ORGANIZATION 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO / PROJECT NO / TASK NO / WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification)

A CLUSTER ANALYSIS OF MANPOWER SUPPLY TO ARMY RESERVE CENTERS

12 PERSONAL AUTHOR(S)
Jones, Alvin R.

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Master's Thesis	FROM ______ TO ______	1987 June	79

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Army Reserve, cluster analysis, classification, supply models, manpower

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

This thesis applies cluster analysis to the problem of grouping Army Reserve Center markets based on measurable economic characteristics of local labor markets and characteristics of individual Reserve Centers. Three applications with potential uses in manpower planning are demonstrated. Predicting models for Reserve accessions are developed for the clustered Reserve Centers.

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL Professors George W. Thomas, Stephen L. Mehay	22b TELEPHONE (Include Area Code) 408-646-2741/2643 22c OFFICE SYMBOL 54Te/54Mp

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

Approved for public release; distribution is unlimited.

A Cluster Analysis of Manpower Supply To Army Reserve Centers

by

Alvin R. Jones
Civilian, U.S. Army Test and Evaluation Command
B.S., University of Baltimore, 1979
M.B.A., University of Baltimore, 1982

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
June 1987

Author: Alvin R. Jones
Alvin R. Jones

Approved by: George W. Thomas
George W. Thomas, Co-Advisor

Stephen L. Mehay
Stephen L. Mehay, Co-Advisor

Willis R. Greer, Jr., Chairman,
Department of Administrative Science

Kneale T. Marshall
Kneale T. Marshall,
Dean of Information and Policy Sciences

ABSTRACT

This thesis applies cluster analysis to the problem of grouping Army Reserve Center markets based on measurable economic characteristics of local labor markets and characteristics of individual Reserve Centers. Three applications with potential uses in manpower planning are demonstrated. Predicting models for Reserve accessions are developed for the clustered Reserve Centers.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION AND LITERATURE REVIEW

A. PROBLEM

Membership in a military organization is unique in many ways. One source of this uniqueness is the overriding importance of the mission of the armed forces: the protection of the nation's vital interests, the deterrence of war, and the attainment of the the nation's objectives by the use of force if war should come. Means to accomplishing the military mission, like many things, can be restricted by budget and manpower considerations. The current budget deficits, together with future obligations connected with a military build-up, are forcing a review of the appropriate levels of reserve and active forces.

To maintain readiness in the face of budget restrictions, military decision makers have been pursuing a policy of increasing reserve manning while maintaining a cap on active force end-strengths. Savings estimates resulting from placing military units in the reserve rather than the active forces are made generally from studies which compare current peacetime costs for existing similar units in the active and reserve forces. These estimates generally show that the saving achieved is a strong function of the type of unit and required readiness or activity level. Units where the capital labor mix is high and where readiness demands high activity levels (more typical of air force and navy flight units) show savings of roughly 25% to 33% for reserve units, whereas more labor intensive units (typical of army infantry units) show savings of as much as 70% [Ref. 1: p. 220].

In addition, planners count on the assumption that reserve forces are less expensive than active forces to maintain because reservists are paid only for the time they actually spend at drills. Also, the contribution that reserve forces make to overall readiness has been increasing steadily since the inception of the of the voluntary force. This is because escalating personnel costs have forced planners to limit the size of active forces, and the removal of the draft has diminished the capability of the active force to quickly expand and mobilize. Currently, any significant mobilization would require reserve augmentation of active forces almost immediately [Ref. 1: p. 8].

To meet this expanding role, reserve forces are organized into three categories: the Ready Reserve, the Standby Reserve, and the Retired Reserve. The Ready Reserve

is the primary contributor to readiness and it is composed of the Selected Reserve and the Individual Ready Reserve (IRR). The IRR consists of individuals who train at irregular intervals and whose role is augmentation of existing units during mobilization. The Selected Reserve is the most significant component of the Reserve force and it consists of units which are organized and equipped to perform specific missions, trained personnel who augment active units, and individuals in training pipelines.

The Selected Reserve of the Department of Defense is made up of six components:

Army National Guard	Marine Corps Reserve
Army Reserve	Air National Guard
Naval Reserve	Air Force Reserve

The Selected Reserve contains combat and support units that would be vital to the successful operation of a major war. For example, the Selected Reserve contains:

* Army: Combat divisions and brigades, armored calvary regiments and numerous support units.
* Navy: Mine warfare ships, amphibious ships and anti-submarine patrol squadrons.
* Marine Corps: A combat division, an air wing and support units.
* Air Force: Fighter, intercepter, tanker and airlift squadrons.

Most members of the Selected Reserve are required to participate in training drills for 24 days a year and in two weeks of annual active duty for training. New enlistees who do not have previous military service also are required to undergo three or more months of initial entry training along with their Active force counterparts. Each member is paid, according to grade, for participating in training.

The impact of recent Defense manpower policy has been that while Active force levels have remained constant over the last decade, Selective Reserve end-strengths have risen from 788,000 in 1978 to 1,100,652 in September 1985 [Ref. 2: p. 1]. A breakdown of current Selected Reserve strength by components is shown in Table 1. This analysis will focus on the Selected Reserve.

Future projections for all components show increases in end strengths for Selected Reserve Forces. For example, the Army manpower plan submitted in the February 1985 budget projected an increase of 116,000 members of the Army Selected Reserve (Army Reserve and Army Reserve National Guard) by 1990. This represents an increase of 16 percent of current end strength over five years [Ref. 3: p. 1].

TABLE 1

SELECTED RESERVE MANPOWER

September 1985

COMPONENT	STRENGTH
Army National Guard	439,952
United States Army Reserve	292,080
United States Navy Reserve	129,832
United States Marine Corps Reserve	41,586
Air National Guard	109,398
United States Air Force Reserve	75,214
DoD Total	1,008,062
United States Coast Guard Reserve	12,590
Total	1,100,652

Source: Defense Manpower Data Center, *Official Guard and Reserve Manpower Strengths and Statistics*, September 1985.

Meeting these expansion requirements efficiently will depend upon a sound understanding of the impact of factors which affect Reserve force supply levels. At present that type of information is not available.

The econometric model is perhaps the most widely used technique for evaluating military personnel supply. Typically econometric manpower supply models attempt to estimate or predict the number of contracts signed by (or actual enlistments of) "high quality" young males based on variables deemed to be related to the enlistment decision. This analysis will explore the use of cluster analysis to classify Army Reserve Centers in relation with local accession factors. These procedures empirically form "clusters" or groups of highly similar entities. Entities involved here are Reserve Centers.

The analysis will explore models estimated for Army Selected Reserve data. This is because Army components represent 67 percent of current Selected Reserve manpower (see Table 1) and Army units are the best examples of units which are forced to survive within the confines of their local labor market. Air Force and Navy Reserve units have more flexibility in recruiting for and manning units from outside their local areas.

For a better understanding of the possible impact that the analysis can have, Table 2 is provided. Table 2 provides a force profile for the DoD Selected reserve in 1985. An analysis of this type can assist policy formulation in the following areas:

* serve as a source of hypothesis about accessions
* allocation of new authorizations across units
* location of new units
* allocation of recruiting resources across geographic areas

TABLE 2

SELECTED RESERVE PROFILE-SEPTEMBER 1985

FORCE	enlisted STR	% MNRTY	% FEM	avg AGE	avg YOS	% HSG
ARNG	397,612	23	5	30	8	57
USAR	238,220	36	17	29	7	50
USNR	106,529	17	11	31	9	64
USMCR	38,204	27	4	24	4	73
ANG	96,361	12	12	34	11	79
USAFR	59,599	23	18	33	10	76
DoD	936,525	25	10	-	-	-

Source: Defense Manpower Data Center, *Official Guard and Reserve Manpower Strengths and Statistics*, September 1985.

B. CLUSTER ANALYSIS

Clustering is the grouping of similar objects. The principle functions of clustering are to name, to display, to summarize, to predict, and to aid in interpretation of data with many dimensions. Clustering techniques were first developed in the field of biological taxonomy. It is one of several methodologies included in the broader category called classification.

The operational objectives of clustering is to classify new observations, that is, recognize them as members of one category or another. This can be contrasted with discriminant analysis where some part of the structure is known and missing

information is estimated from labeled samples. In cluster analysis little or nothing is known about the category structure. All that is available is a collection of observations whose category membership are known (variables). The analysis seeks to discover a category structure which fits the observations. The problem may be stated as one of finding the natural groups, which means to sort the observations into groups such that the degree of "natural association" is high among members of the same group and low between members of different groups.

Most of the well known clustering techniques fall into one of two main categories: (1) hierarchical and (2) nonhierarchical (partitioning). [Ref. 4: p. 124]. The former is one in which every cluster obtained at any stage is a merger of clusters at previous stages. The nonhierachial procedures however form new clusters by lumping and splitting old ones.

In a geometric sense, every observation may be viewed as a point in p-dimensional euclidean space. [Ref. 4: p. 127]. This swarm of data points may contain dense regions or clouds of data points which are separable from other regions containing a low density of points. These denser regions constitute what are known as clusters. In one and and two dimensional cases, it is easy to visualize and detect clusters from scatter plots, assuming that clusters exist. In higher dimensions (which will be used in this analysis) clustering becomes extremely difficult without the aid of a computer.

Cluster analysis techniques have been applied in many fields of study. The terminology differing from one field to another in literature is both voluminous and diverse. "Numerical taxonomy" is frequently substituted for cluster analysis among biologist, botanist, and ecologist, while some social scientist may refer to "typology." Other frequently encountered terms are pattern recognition and partitioning. While techniques such as discriminant analysis have been studied by statisticians for nearly 45 years, cluster analysis has only recently come to statistical notice. Any method which partitions a set of objects into subsets on the basis of measurements taken on every object qualifies as a clustering method.

C. REVIEW OF LITERATURE

1. Active Force Supply Studies

The purpose of this review is to present variables which have been found to be important in military accession research, and provide a better understanding of cluster analysis. Understanding the importance of the variables in conjunction with cluster

analysis is critical in any eventual application of findings to the management of military forces.

There is one problem which continually recurs in studies of military supply. Almost all studies use regression analysis to model supply levels. Because of data constraints the dependent variable used is a measure of enlistment contracts signed or accessions. The problem is that the variable which is being modelled, potential military supply, is not always the same as the number of recruits enlisted. This is because the service set quotas on enlistment levels. These quotas vary between services. Also, within services, different quotas apply for different categories of recruits. This means that the variable researchers are measuring, supply, is actually a function of both potential supply and enlistment quota. The implication is that results of studies which use demand constrained data do not accurately reflect the underlying relationships between the economic environment and potential supply. Methods which have been used to overcome this problem are discussed in more detail below.

In a May 1985 study, Dertouzos pointed out that previous studies of factors influencing the supply of enlistments did not consider the effects of demand, such as the enlistment goals and incentives that are set up for recruiters to secure high quality recruits. His analysis demonstrates that that enlistments are produced through the simultaneous interaction of both supply and demand factors [Ref. 5: p. 3]. This suggests that past research results that ignore demand are likely to have been flawed by significant estimation biases. That is, changes in such factors as unemployment, relative wage rates, and recruiting resources can affect enlistments more than past studies have indicated.

Lawrence Goldberg conducted a comprehensive study which developed an econometric supply model for all services using pooled time series, cross section recruiting data from 1976 to 1980. The model was developed using log linear ordinary least squares regression. The dependent variable in the model was the number of male nonprior service (NPS) high school graduates (HSG). The model was estimated separately for all HSG and those in mental categories I-IIIA. The dependent variables in the model were:

* relative military / civilian pay
* civilian unemployment
* military education benefits
* expenditures on Federal youth employment programs

* race

* number of recruiters by service

* Navy advertising budget (other services data were unavailable)

Goldberg handled the problem of demand constraints by focusing his analysis on the results pertaining to the male high quality sample (i.e. mental category I-IIIA HSG). This is a standard procedure in supply modelling. He claimed that this group is rarely demand constrained and therefore his model should produce accurate results [Ref. 6: p. 10].

A study by Daula and Smith rejected Goldberg's contention that using high quality enlistee samples removes the problem of demand constraint contamination in study results. They contend that even high quality groups may be demand-constrained in certain geographic areas. [Ref. 7: p. 6] To overcome this problem they partition their data into two samples. One sample is data from areas where recruiting goals are met (i.e. supply constrained). The other sample is all the demand constrained data. The total sample consists of time series, cross section data from 54 Army recruiting districts by month from October 1980 to June 1983.

They included the following independent variables in a log-linear OLS regression:

* military pay and bonuses

* civilian pay

* unemployment

* qualified military population

* percent minority

* percent voting Republican

* enlistment goals for all services

* number of Army recruiters

* levels of national and regional advertising.

One important result from this study came from estimating the supply function using only supply-constrained data but including the high quality enlistment goal as an independent variable. The resulting coefficient of the goal variable was not significantly different from zero, indicating that recruiters goals have no effect on

enlistments in supply-constrained districts. This result supports the validity of Daula and Smith's data partitioning methodology.

In a subsequent study, Goldberg and Greenston reported on the updating and further development of a basic time series cross section model, which analyzes the supply of nonprior service, male, upper mental category enlistments. This study updates the data base to include FY 1983 observations, develops better measures of key variables (civilian pay, unemployment, and population), and reestimates the model with data for the longer period FY 1976-1983. [Ref. 8: p. 61] A major improvement was in the use of unemployment data. They used an annual measure for "each" Navy recruiting district (NRD) based on the aggregation of monthly county-level data from the Bureau of Labor Statistics. A framework of the model is the contention that enlistment is viewed as an employment decision that is heavily influenced by economic considerations. It is assumed that an individual compares two employment alternatives--work in the military or work in the private sector--and chooses the one that maximizes economic benefits. This implies that the enlistment propensity will increase if there is an increase in the economic benefit of working in the military (such as an increase in military pay) or a decline in economic benefits of working in the private sector (such as an increase in unemployment). In addition, the authors assume that enlistment supply in a NRD depends on the enlistment propensity of the districts residents and on the number who are eligible for enlistment.

In the Goldberg and Greenston study, propensity and eligibility are influenced by various controllable and exogenous factors, which are grouped into broad categories: economic and demographic factors, recruiting resources, and policies. The economic factors include relative military pay, civilian unemployment, and GI Bill benefits. The demographic factors are the civilian male youth HSG and high school seniors population, racial mix, and urban/rural mix. The recruiting resources are recruiters of each service, and the recruiting policies considered are Air Force and Marine Corps changes in goals and standards. No additional consideration was given to problem of demand constraints.

Results of active force supply studies are not directly applicable to the Reserves for several reasons. Among the reasons is the fact that the majority of reservists have a full time civilian job and participation in the reserves is a moonlighting decision. Another reason is the Active force recruits and operates in a national labor market. The Reserves, particularly the Army components, are forced to

operate in local labor markets. Across the U.S. the economic and demographic factors that affect enlistments vary considerably among local areas.

A restriction to consideration at the local labor market level does not negate the importance of demand constraint in the Reserve supply modelling process. It is still plausible that in local areas potential supply may exceed recruiting quotas. A further complication is that the impact of quotas will be different across local labor markets because of the differences in the magnitudes of factors affecting potential supply.

2. **Reserve Supply**

There have been very few studies on Reserve enlistment supply since the introduction of the volunteer force. A few of them are discussed in the following paragraphs.

Kelly, in 1979, estimated supply models for both NPS and PS personnel using total DoD accessions as his dependent variable and relative pay, unemployment and population as independent variables. This analysis is disaggregated to the state level and he derived relative wage elasticities of .35 for PS supply and .10 for NPS supply [Ref. 9: p. 1].

As a part of a large study to investigate the impact of the all-voluntary-force on the Air Force Reserve, Rostker developed the moonlighting model. The model characterizes the choice to work as a tradeoff between the individual's desire for income (from work) and leisure time [Ref. 10: p. 299]. In two subsequent studies, McNaught, reviews the work of Rostker and Kelly and points out a number of limitations and inconsistencies in their results. Combining those studies and the moonlighting model, McNaught [Ref. 11: p. 12] conceptualized a theoretical model of reserve supply where:

R = f(W, C, S, H, U, P, I, T, X)

R = measure of reserve participation

W = Reserve wage

C = civilian primary wage

S = civilian secondary wage

H = hours worked on primary job

U = unemployment rate

P = population of eligible enlistees

I = stock of available information about Reserves

T = travel costs

X = a set of regional variables

McNaught's final model recognizes the significance of available information about Reserve enlistment opportunities as a determinant of Reserve participation. Because of data restrictions, the model which McNaught estimates is much more restrictive than his theoretical model. Specifically, he disaggregates his data to the state level and includes no measure of travel cost (which is not reimburse able), Reserve opportunity information (availability of better jobs), or recruiting goals. In his estimation McNaught concentrates on NPS enlistments and he looks at total DoD accessions without a separation by component. He estimates his model using logit analysis with the ratio of number of prior service accessions to qualified population as the dependent variable. This specification attempts to predict the probability of an individual with a given set of characteristics enlisting in the Reserves [Ref. 11: p. 36].

Borack et. al (1985) list four criticisms of McNaught's study:

(1) level of aggregation was too high

(2) lack of measure of regional military interest

(3) no consideration of the interaction between Reserve and Active recruiting systems

(4) no consideration of the effect of local recruiting goals (demand) on enlistment supply by geographical area.

[Ref. 12: p. 36]

Grissmer and Kirby in an effort to help fill the gap in research on Reserves, analyzed the attrition and reenlistment decisions of NPS, enlisted personnel in the Army Reserve and Army National Guard [Ref. 13: p. 130]. They point out that reserve participation resembles civilian moonlighting in some respects, but there are also some major differences:

(1) Reservists are legally commited to their term of service;

(2) All reservists must leave their primary job for at least two weeks annually to work full time on the Reserve job, and new nonprior service reservists must additionally train full time for at least four months;

(3) Reservist drill a limited, specified number of hours and therefore do not have the option of working more to earn more;

(4) The Reserve job offers nonpecuniary benefits such as specialized training, as well as an environment that may generate camaraderie and a sense of team accomplishment, and finally;

(5) Reservists receive other fringe benefits of military service such as educational benefits and exchange privileges (while on annual training).

The reserve supply studies studies reviewed above are inconsistent and of limited use in estimating the effect of policy and demographic changes on potential supply. To improve the models the following considerations should be incorporated:

Data should be analyzed at the lowest level possible (local Reserve Centers).

The impact of recruitment goals and quotas should be included.

Accessions should be modelled by individual Reserve component.

Cross effects of own and other service, Active and reserve recruiters should be included.

Theoretical analysis of the Reserve participation decision and a review of previous military supply studies suggest that a useful model of Reserve supply should explain the number of Reserve component accessions as a function of the following explanatory variables:

Economic Factors

local unemployment rates

Reserve compensation

Reserve benefits

civilian primary job wages

hours worked on primary job

Demographic Factors

age

race

education

family incomes

family sizes

distances to Reserve Centers

Recruitment Policies

recruitment goals

recruiters by component

local military interest

advertising effort

Restrictions on available data may limit the use of suggested variables in the subsequent analysis.

3. Cluster Studies

Although classification is a fundamental step in the process of scientific studies, different sciences have different problems to solve. In addition, classification often contains the concepts necessary for the development of theories within a science.

"Cluster analysis" is the generic name for a wide variety of procedures that can be used to create a classification. These procedures empirically form clusters or groups of highly similar entities. More specifically, a clustering method is a multivariate statistical procedure that starts with a data set containing information about a sample of entities and attempts to reorganize these entities into relatively homogeneous groups. [Ref. 14: p. 7] Clustering methods have been recognized throughout this century, but most of the literature on cluster analysis has been written only during the past two decades. Cluster analysis has taken many forms and is often defined in many, sometimes contradictory, ways. Literature on cluster analysis can be found in a variety of journals, ranging from electrical engineering to biology to library science to psychiatry.

The major stimulus for the development of clustering methods was a book entitled *Principles of Numerical Taxonomy*, published in 1963 by two biologists, Robert Sokal and Peter Sneath. Sokal and Sneath argued that an efficient procedure for the generation of biological classifications would be to gather all possible data on a set of organisms of interest, estimate the degree of similarity among these organisms, and use a clustering method to place relatively similar organisms into the same groups. [Ref. 14: p. 9, citing Sokal and Sneath, 1963]. Once groups of similar organisms were found, the membership of each group could be analyzed to determine if they represented different biological species. In effect, Sokal and Sneath assumed that "pattern presented process"; that is, the patterns of observed differences and similarities among organisms could be used as a basis for understanding the evolutionary process. The literature on cluster analysis exploded after the publication of the Sokal and Sneath book.

Solomon [Ref. 15: p. 37] lists three major avenues of approach in solving a clustering problem:

(1) Total enumeration of all data partitions and the subsequent selection of a good or optimal clustering configuration;

(2) A stepwise clustering scheme that selects for each number of clusters the best available groupings with the realization that it may ignore some good configurations in the process;

(3) Reduction of multivariate data to two or three orthogonal dimensions, producing a graphic or pictureal representation that permits visual clustering.

An essential step in any of these approaches is representation of the data and establishment of measures of similarity. Since the choice of the variables to be studied, their interrelationships and the measures of similarity are the basis for any clustering scheme, much consideration must be given to ensure that "closeness" in the sense of the similarity measures indicates closeness in the sense of the objectives of a study. The simplest and most common measures of similarity are those which combine the effects of individual variables into a single number. This assumption of numerical comparability allows clustering processes that group objects by overall similarity. Ball [Ref. 16: p. 17] lists five types of similarity measures:

(1) Association: The similarity between object X and object Y is the number or a function of the number of variables for which X and Y have the same response;

(2) Correlation: Correlation between object X and object Y is a function of the angle between their respective vectors; it is most useful when a pattern of ratios of the variables is the prime determinant of similarity;

(3) Distance: Many different distance measures are available. Weightings can be applied to absolute or euclidean distances and can be derived either from an a priori evaluation of each variable's importance or from the data. Euclidean distances were emphasized by Ball;

(4) Probabilistic: These measures are used primarily when it is appropriate to modify weights of the variables on the basis of population statistics;

(5) Functional: For functional measures, the value of similarity is a function of the distance from other objects.

When measures of similarity between objects have been established, the measures must be modified to provide meaningful similarity between groups of objects and between objects and groups.

Alexander, in 1974, examined the relationships between the structure of internal labor markets and the mobility, experience and income of workers. [Ref. 17: p. 64] In order to examine the relationships between structure and variables, he realized a measure was required that would allow him to classify internal labor markets. Most

of the previous research in this area has been based on case studies in which an industry has been intensively analyzed and subjectively classified. One of the goals of his work was to develop classification criteria of structures, based on objective and comprehensive data, that were consistent with the results of case studies. Internal labor markets were classified according to many different schemes, but he utilized Kerr's taxonomy of open, guild, and manorial markets. The open market is the unstructured, competitive type. Guild-type markets are stratified horizontally. And manorial markets emphasize attachment to the place of work and vertical stratification. Alexander concluded that segmentation does exist because of institutional characteristics.

Milligan, in 1980, conducted an evaluation of several clustering methods. [Ref. 18: p. 325] He acknowledged that a general definition of cluster structure was unlikely, but he offered one which involves two parts. Essentially, clusters should exhibit the properties of external isolation and internal cohesion. External isolation requires that entities in one cluster should be separated from entities in another cluster by fairly empty areas of space. Internal cohesion requires that entities within the same cluster should be similar to each other, at least within the local metric. This definition is similar to the concept of natural clusters.

To evaluate 15 clustering methods, he created a data set which would naturally cluster. Then in conjunction with the clustering methods, he added, one at a time, six different error perturbations. These were:

(1) Error-free parent data sets.

(2) Data sets with outliers.

(3) Error perturbation of the distances.

(4) Addition of random noise dimensions.

(5) Computation of distances with noneuclidean index.

(6) Standardization of the variables.

The simulated data sets were clustered by eleven agglomerative hierarchical algorithms and four nonhierarchical centroid sorting procedures. The methods are listed in Table 3. The last four methods are nonhierarchical (k-means) centroid sorting procedures which produce only a single partition.

The set of methods was chosen primarily for three reasons. First, program listings for the methods are generally available and can be adopted for many types of

TABLE 3

HIERARCHICAL METHODS AND K-MEANS ALGORITHMS

METHOD	ERROR FREE	HIGH ERROR
Single Link	.974	.777
Complete Link	.995	.880
Group Average	.998	.948
Weighted Average	.994	.934
Centroid Method	.983	.810
Median Method	.976	.831
Ward's Minimum Variance	.987	.940
Beta-Flexible	.997	.945
Average Link in the New Cluster	.985	.906
Minimum Total Sum of Squares	.935	.835
Minimum Average Sum of Squares	.993	.919
MacQueen's Method	.884	.842
Forgy's Method	.932	.872
Jancey's Method	.927	.909
Convergent K-means	.903	.897

Source: Milligan, G.,
An Examination of the Effect of Six Types of Error Perturbation on Fifteen Clustering Algorithms
Psychometrika-vol. 45, no. 3. September, 1980, p. 332.

clustering problems. Secondly, the methods are all fairly fast in terms of CPU time and are economical for most applications. Further, some of the methods have been adapted to handle very large data sets. He concluded that the results indicated the hierarchical methods were differently sensitive to the type of error perturbation. Also, he indicated that the simulation results were promising and a more detailed study of this and other such indices should be undertaken.

Hodson in 1983, [Ref. 19: p. 25] employed a rigorous approach to defining market sectors. He began with data on 40 characteristics of firms or industries, encompassing firm size, productivity, unionization and various market measures, such

as government regulation and foreign involvement. Principle components analysis was used to reduce the 40 variables to 25 factors. He then applied cluster analysis with the factor scores from 202 industries to form 16 industry clusters. He collapsed the 16 clusters into six industry groups to facilitate an empirical analysis of earnings. Rather than use a clustering algorithm, the final grouping was based on the authors judgement. He criticized previous work on dual labor markets for relying on only two groups. Hodson found that industry group affected earnings, even when worker characteristics were held constant. His findings were inconsistent with other work by sociologists and raised many questions on labor market theory.

As demonstrated in the review of cluster analysis there is a diversity of disciplines contributing to the literature. There is also a variety of methods lumped under the term cluster analysis. This thesis will pursue methods relevant to the Reserves, incorporating known econometric techniques and cluster analysis.

II. DATA/METHODOLOGY

A. DATA

The primary research question is to identify specific social and demographic factors among local geographic areas that can provide a basis for classifying Army Reserve Center markets into unique and homogeneous groups. Classifying characteristics are social and demographic factors related to the local labor market and recruiting success measures attributable to units attached to the reserve center.

To conduct this study, data were extracted from the mass storage volume group at the U.S. Army Recruiting Command (USAREC). The file contains reserve accession counts and other accession variables for use in analysis. Another data file composed of demographic and local labor market factors has been created at the Naval Postgraduate School. This file was merged with the USAREC file to match accessions with local market data. The merger gave a final file which contained 967 records.

Each record contains accession counts, occupation and industry counts, black population percentages, unemployment figures, income, family size, rental home value information, recruiter counts, authorization data, military available figures, member and unit strength data, and wage data. Unemployment, wage figures, and accession counts are from 1985. This matched file will be the basis for the similarity analysis conducted in this study. Table 4 identifies summary statistics for nonprior service and prior service male and female reserve accessions. Later, in an effort to reduce the sample size, a random sample within the range identified will be utilized. Accessions for all reserve components totaled 306,108. Observations or cases are Reserve Centers.

For reading convenience, a sample of the range in variable values (minimum and maximum values) are shown in Table 5. A further description of these and other variables are shown in Appendix A-C. Each market is defined by local factors within a 35 mile radius. Because the reserve markets are defined geographically by distance, there tends to be a wide disparity in characteristics. The file contains data on reserve services other than Army, but this analysis will concentrate on a subset of data representing accessions to the Army Reserves. Local labor market conditions are more likely to have a greater impact on Army Reserve Supply due to the large number of USAR Reserve Centers.

TABLE 4

1985 ACCESSION DATA

ANPSAA	ARMY NONPRIOR SERVICE MALE & FEMALE ALL AGES		
ANPSAA	231,250.0		
MEAN	239.1	STD DEV	339.7
MINIMUM	0	MAXIMUM	1903.0
APSAA	ARMY PRIOR SERVICE MALE & FEMALE ALL AGES		
APSAA	74858.0		
MEAN	77.4	STD DEV	113.8
MINIMUM	0	MAXIMUM	632.0
VALID OBSERVATIONS - 967		MISSING OBSERVATIONS - 0	

B. METHODOLOGY

A focus of this thesis is to identify local labor market variables useful in grouping relatively homogeneous groups of Reserve Centers. The cluster analysis involves multivariate statistical procedures. The heart of any multivariate analysis consists of the data matrix (Table 6). This matrix is a table that gives a number of observations on a number of variables simultaneously. For this study, observations are Reserve Centers and variables are local labor market characteristics which effect supply and characteristics of the Reserve Center. The following is a discussion of cluster analysis methodology.

Clustering methods are used to discover structure in data that is not readily apparent by visual inspection or by appeal to other authority. The analysis is a two stage process. The first stage is to choose quantifiable attributes that describe the objects, and then use these attributes to measure the pair-wise dissimilarity among the objects. The second stage is to represent these dissimilarities by an appropriate classifying system or display.

The input to cluster analysis is normally an n x p matrix of data. Measurements of p attributes for each of n objects. In this case it will be measurements of variable

TABLE 5

LABOR MARKET SUMMARY DATA

Variable	Minimum	Maximum
NPS Male	0	1,511
QMA Male	1	11,384
% of Pop Black	0	100
% Wt M unemp	0	19
% Blk M unemp	0	18
Ave Fam incm	9,849	30,127
Ave Family size	3	5
Med Home Val	19,556	127,966
Med Home Rnt	98	368
% Fam 2 Wrkrs	0	100
%Pop Chg70-80	0	100
% Wkrs by Industry		
Manufacturing	0	100
Service	0	100
Government	0	100
Seasonal	0	100

characteristics for each Reserve Center local area. The output from cluster analysis [Ref. 20: p. 47] is normally one of three displays:

A hierarchical classification, commonly called a tree diagram or dendrogram;

A partition of the objects into mutually exclusive sets, each set described by a profile or vector of p attribute values;

A clumping of objects into sets that may overlap, each set again described by a profile.

In particular, the output should highlight mutual interaction among three variables or more, just as easily as one can highlight a two way interaction. The value of these outputs is that they summarize the original data objectively and they tend to

TABLE 6

ILLUSTRATIVE DATA MATRIX

	Variables				
Observations	1	2	3.....	j.....	p
1	X11	X12	X13	X1j	X1p
2	X21	X22	X23	X2j	X2p
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
i	Xi1	Xi2	Xi3	Xi4	Xip
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
n	Xn1	Xn2	Xn3	Xn4	Xn5

highlight subtle interactions in the original data, enabling a user to formulate reasonable hypotheses about the interactions.

Things that are recognized as similar or dissimilar are fundamental to the process of classification. [Ref. 21: p. 13] Despite its apparent simplicity, the concept of similarity, and especially the procedures used to measure similarity, are far from simple. Similarity does not lie with the simple recognition that things are either alike or not alike, but instead in the ways in which these concepts are expressed and implemented in scientific research. To be successful, research has to be based upon objective procedures. Cluster analysis is a result of this necessity.

Often the term "similarity coefficient" (or measure) is used to describe any type of similarity measure. Sneath and Sokal (1973), subdivided these coefficients into four groups:

(1) correlation coefficients,

(2) distance measures,

(3) association coefficients, and

(4) probabilistic similarity measures.

For this analysis, distance measures will be used.

The quantitative estimation of similarity has been dominated by the concept of metrics. Any nonnegative real valued function d(x,y) can be used to judge whether a similarity measure is a true distance function (or metric).

(1) Symmetry. Given two entities, x and y, the distance, d, between them satisfies the expression

$$d(x,y) \geq 0$$

(2) Triangle inequality. Given three entities, x,y,z, the distances between them satisfies the expression

$$d(x,y) \leq d(x,y) + d(y,z)$$

This simply states that the length of any side of a triangle is equal to or less than the sum of the other two sides. This concept has also been called the metric inequality.

(3) Distinguishability of nonidenticals. Given two entities x and y,

$$\text{if } d(x,y) \neq 0, \text{ then } x \neq y$$

(4) Indistinguishability of identicals. For two identical elements, x and x′

$$d(x,x') = 0$$

The distance between the two entities is zero.

Because of their intuitive appeal, distance measures have enjoyed widespread popularity. Technically, they are best described as dissimilarity measures; most of the more popular coefficients demonstrate similarity by high values within their ranges, but distance measures are scaled in the reverse. Two cases are identical if each one is described by variables with the same magnitudes. In this case, the distance between them is zero. Distance measures normally have no upper bounds, and are scale dependent. The most commonly used distance is the Euclidean distance. It is defined as:

$$\text{distance}(x,y) = \text{SQRT}(\text{sum}(xi\text{-}yi))^2$$

The potential user of cluster analysis should be aware that many types of similarity exist, and that while many of the coefficients and measures commonly used in quantitative approaches to classification are metrics, there are alternatives to the use of these measures that may be appropriate and necessary within the context of research. Choosing a distance function is no less important than the choice of

variables to be used in the study. The choice of similarity measure should be embedded ultimately within the design of research, which is itself determined by the theoretical, practical, and philosophical context of the classification problem. A Euclidean distance measure will be used in this study.

Variable selection to be used with cluster analysis is one of the critical steps in the research process. Ideally, variables should be chosen within the context of an explicitly stated theory that is used to support the classification. The theory is the basis for the rational choice of variables to be used in the study. Traditional theories of labor market participation and the military manpower supply research previously undertaken provide a starting point for identification of variables. From the literature review, various economic, demographic, and recruitment variables are listed. Table 5 lists several candidate variables which could be used in cluster analysis. These variables include; average family income, percent black population, and civilian jobs in the area.

It will be appropriate to standardize all of the variables used in this cluster analysis. In most statistical analysis, the data are routinely standardized by some appropriate method. If the normality of a variable is in question, a logarithmic or other transformation is often performed. If the data are not of the same scale values, they are commonly standardized to a mean of 0 and to unit variance. There is some controversy as to whether standardization should be a routine procedure in cluster analysis. Most of the literature argues convincingly that standardization is inappropriate when the difference in scale between two variables may be intrinsic; but no intrinsic differences seemed likely in the candidate variables used here. Users with substantially different units of measurement will undoubtedly want to standardize them, especially if a similarity such as Euclidean distance is to be used. The decision to standardize should be made on a problem to problem basis, and users should be aware that results differ solely on the basis of this factor, although the magnitude of the effect will vary from data set to data set. Using unstandardized Euclidean Distance in the current situation would clearly result in the dissimilarity coefficient being driven by median home value, and average family income while variables such as average family size would be ignored. Standardization also puts all the variables in comparable units. Each variable used in this analysis will be transformed to a Z-score variable. The Z-score variable transformation standardizes variables with different observed scales to the same scale.

Other types of data transformation are possible, and many of these have been used concurrently with cluster analysis. Factor analysis or principle components analysis is often used when a researcher knows that the variables in the study are highly correlated. The uncritical use of highly correlated variables to compute a measure of similarity is essentially an implicit weighting of these variables. That is, if three highly correlated variables are used the effect is the same as using only one variable that a weight three times greater than any other variable.

The data file used in this analysis contains at least 40 variables. Value listings are included in Appendix D. To make efficient use of the candidate variables, a factor analysis was run on all the variables at once. This analysis will utilize one representative variable from the various groups. Also, the number of Reserve Centers (967) in the data base will be scaled down to explore clusters associated with high accessions, low accessions, fill rates, take, and relative take (compared with National Guard accessions). At the same time, this will have the effect of reducing the data set to a more manageable size.

The SPSSX information analysis system is a comprehensive tool for managing, analyzing, and displaying information. Its capabilities include hierarchical and nonhierarchical techniques. Hierarchical agglomerative methods have been dominant among the seven families of methods in terms of frequency of their applied use. In the agglomerative methods, you begin with N clusters; i.e., each observation constitutes its own cluster. In successive steps the two closest clusters are combined, thus reducing the number of clusters by one in each step.

The K-means clustering is a popular nonhierarchical clustering technique. For a specified number of clusters K the basic algorithm proceeds in the following steps:

(1) Divide the data into K initial clusters. The members of these clusters may be specified by the user or may be selected by the program, according to a predetermined procedure;

(2) Calculate the means or centroids of each of the K clusters;

(3) For a given case, calculate its distance to each centroid. If the case is closest to the centroid of its own cluster, leave it in that cluster; otherwise, reassign it to the cluster whose centroid is closest to it;

(4) Repeat step 3 for each case·

(5) Repeat steps 2, 3, and 4 until no cases are reassigned;

For this analysis, the K-means clustering (nonhierarchical) technique will be used. This is chosen, in part, to handle a large number of cases and a specified number of clusters will be requested. In addition, hierarchical clustering will be used on a smaller sample of the data set.

Four factors appear to influence greatly the performance of clustering methods:

(1) elements of cluster structure,

(2) the presence of outliers and the degree of coverage required,

(3) the degree of cluster overlap, and

(4) choice of similarity measure.

[Ref. 22: p. 23].

To review the cluster methodology, a considered first step is to selectively reduce the size of the data file. What results is a set of variables relevant to the reserves with local characteristics. Next, is a choice of dissimilarity coefficients, and finally the choice of a clustering algorithm.

III. CLUSTER RESULTS

A. CLASSIFICATION

If manpower supply researchers were to try to classify Army Reserve markets,they would probably link demographic, economic and recruiter factors in the decision. These measurable factors would then be used to form a mathematical equation to predict and classify such things as accession counts, fill rates, and relative accessions.

Cluster analysis has more potential as a factor in classification transformation. First of all, the ability to group two Reserve Centers together is intrinsic to every clustering algorithm (so long as the complete link-furthest neighbor sorting strategy is not used). Secondly, cluster analysis requires the user to define only a transformation from measurable factors to a pair-wise dissimilarity coefficient rather than a transformation from measurable factors to dependent measures.

B. ACCESSION COUNTS

1. Nonhierarchical

To demonstrate this application of cluster analysis, the following local area factors were selected (utilizing theories from previous researchers) with which to objectively classify Reserve Center markets into homogeneous groups by accession counts:

1. percentage manufacturing industry
2. average family size
3. primary male military available 17-21 years old
4. unemployment - black male
5. population change 1970-1980
6. mean civilian wages
7. military installations count.

Sample values of these and other variables are listed in Appendix E. Using the variables above, the data set was converted into clusters. Cluster results with measurement characteristics are summarized in Table 7.

The Reserve Center local market data have been clustered with the K-means procedure of Quick Cluster in SPSSX. This method demonstrates the basic features of

TABLE 7

ACCESSION CLUSTER RESULTS

FINAL CLUSTER CENTERS

CLUSTER	ZUNEMPB	ZAVGFAMS	ZMANUFAC	ZWAGES
1	.612	.076	.070	.970
2	.782	2.706	-.756	-.577
3	-.202	-.205	.018	-.154

CLUSTER	ZPOPCHNG	ZMILINS	ZPMILAVA
1	-.926	1.772	1.704
2	.878	-.305	-.407
3	.140	-.346	-.329

NUMBER OF CASES IN EACH CLUSTER.

CLUSTER	CASES	MEAN NPS ACCESSIONS
1	146	853
2	48	93
3	742	122
MISSING	31	
TOTAL	936	

nonhierarchical clustering methods. In the first step, preliminary calculations are made, such as the variable means and standard deviations. Then, an initial partition of the data is obtained with an internally generated starting partition, assigning the 967 market areas into three clusters (K = 3). The next step forms the initial cluster centers. Each of the other observations is assigned to the nearest cluster. Euclidean distance is used for this initial phase, and the cluster centroids are recomputed after each observation is assigned to a group.

After the initial solution has been found, the program advances to the iterative K-means phase. The distance from each observation to each cluster centroid is again computed, using the Euclidean distance criterion, and the assignment to the closest centroid is made and the centroid updated to reflect its new membership. After considering all observations in this manner, the new criterion value is checked for possible improvement during the K-means iteration. As long as the criterion value improves, the K-means procedure is repeated until final cluster centers are found. The final cluster centers in Table 7 result from the variable means for the cases in the final clusters.

Classification of Reserve Centers by local factors should depend primarily on pair-wise data. The factor data were standardized into Z-values to get a meaningful data set. Appendix E lists the variable transformations, and Appendix F lists the Z-values.

As shown in Table 7, the three cluster iteration results in separate clusters of 146, 48, and 742 local market areas. Table 7 also shows the mean nonprior service accession counts ranging from 93 to 853 for each of the clusters.

This suggests that clusters according to accession counts can be classified as:

* 146 high accession market areas (cluster 1);
* 742 medium accession market areas (cluster 3) and;
* 48 low market areas (cluster 2).

There appears to be a significant difference in the average accessions from high market areas, as opposed to accessions from the medium or low areas. A further investigation of cluster membership reveals that for the high accession market areas, none of the 146 cases have accessions lower than 133, and the highest accession count for this cluster is 1903. Clusters two and three are not as distinct in grouping median and lower clusters, in that the range is from 6 to 361 for cluster 2, and 0 to 730 for cluster 3. (See Appendix G for cluster statistics). This could indicate the existence of outliers in each of the clusters.

A natural question to ask after observing the results of a cluster analysis is what variables most strongly influence the clustering observed. A clue could be provided by a look at the mean and standard deviations of the cluster member variables. Table 8 shows that cluster 1 is distinguished from the other clusters with high average values in primary military available, mean wages, and military installations count. And a low value for population change. Primary military available is almost ten times the average of the other clusters, as well as Military installations count. It should be noted that these values also correspond (relatively) to values shown for final cluster centers in Table 7.

One major problem shared by all iterative methods is the problem of suboptimal solutions. Since these methods can sample only a very small proportion of all possible partitions of a data set, there is some possibility that a suboptimal partition may be chosen. Unfortunately, there is really no objective way to determine if a solution from an iterative partitioning method is globally optimal. One avenue of solution to the problem, however, is to use the clustering method in conjunction with

TABLE 8

CLUSTER CHARACTERISTICS

	CLUSTER 1 N = 146		CLUSTER 2 N = 48		CLUSTER 3 N = 742	
VARIABLE	MEAN	STANDARD DEVIATION	MEAN	STANDARD DEVIATION	MEAN	STANDARD DEVIATION
MANUFACT	.3	.06	.2	.06	.3	.09
AVGFAMSZ	3.4	.07	3.8	.19	3.3	.13
PMILAVAL	205.1	142.31	16.9	12.20	23.9	22.85
UNEMPB	6.6	1.59	7.2	5.12	4.0	3.06
POPCHNGE	0.0	.06	.3	.17	.2	.14
WAGES	9.3	1.32	7.0	1.07	7.7	1.32
MILINS	40.2	22.39	4.6	6.35	4.0	6.63

an appropriate validation procedure. One validation procedure could be the use of regression analysis.

2. Regression

Multiple regression, the use of many independent variables to predict a dependent variable, is probably the statistical technique used and understood most often by managers. An attempt will be made here to develop a predicting equation for the clustered nonprior service accession counts using the previous seven variables (as independent or explanatory variables). Nonprior service accessions are used as the dependent variable.

Results obtained from estimating the multiple regression model of nonprior service accessions for all of the Reserve Center markets are shown in Table 9. All of the variables are statistically significant at the 5% level. The variables, primary military available, population change, mean wages, and military installations count are significant at the 1% level. This equation would suggest, in simple terms, that high accessions would be found in market areas where the percentage manufacturing industry, average family size, primary military available, mean wages, and military installations count are relatively high. And where black male unemployment and population change are low.

Accession equations were estimated for each of the three clusters and shown in Tables 10, 11, and 12. These results indicate that high accession market areas

TABLE 9

NPS ACCESSION MODEL

EXPLANATORY VARIABLES	COEFF	t	PROB VALUE
PERCENTAGE MANUFACTURING INDUSTRY ...	112.4	2.5	.012
AVERAGE FAMILY SIZE	49.8	2.4	.016
PRIMARY MALE MIL AVAILABLE 17-21/10	2.9	47.0	.001
UNEMPLOYMENT-BLACK MALE	-2.3	-2.0	.045
POPULATION CHANGE 1970-1980	-220.7	-8.3	.001
MEAN WAGES	11.0	4.2	.001
MILITARY INSTALLATIONS COUNT	3.6	11.4	.001
INTERCEPT TERM	197.6	-2.7	.010

Dependent variable-Nonprior Service Accession Counts

N = 967

R SQUARE = .909

(cluster 1) are significantly influenced by the variables primary military available, mean wages, population change, and military installations count. Percentage manufacturing industry, unemployment black male, and average family size are not significant for this cluster. In addition, the sign of the coefficient for average family size changed from positive to negative.

Medium accession market areas (cluster 3) are influenced significantly by primary military available, average family size, black male unemployment, population change, and military installations count. Percentage manufacturing industry is not statistically significant and mean wages is not significant at the 5% level. For this cluster, signs changed from positive to negative on coefficients for variables of percentage manufacturing industry mean wages, and military installations count. This would imply that medium accession market areas would cluster where these factors are low.

Table 11 shows that medium to low accession market areas (cluster 2) are influenced significantly (1% level) by primary military available alone. Military installation counts and black male unemployment are significant at the 10% level. For this cluster, population change, mean wages, percentage manufacturing industry, and average family size do not significantly influence the equation. Results for this equation may be affected by a small sample (48).

TABLE 10

HIGH ACCESSION REGRESSION MODEL

EXPLANATORY VARIABLES	COEFF	t	PROB VALUE
PERCENTAGE MANUFACTURING INDUSTRY ..	2.3	15.8	.001
AVERAGE FAMILY SIZE	-170.4	0.8	.411
PRIMARY MALE MIL AVAILABLE 17-21/10	2.3	15.8	.001
UNEMPLOYMENT-BLACK MALE	-13.2	-1.4	.169
POPULATION CHANGE 1970 - 1980	-207.7	-8.1	.001
MEAN WAGES	32.9	2.7	.007
MILITARY INSTALLATIONS COUNT	5.2	3.9	.001
INTERCEPT TERM	391.9	0.5	.618

Dependent variable-Nonprior Service Accession Counts

N = 146

R SQUARE = .878

TABLE 11

LOW ACCESSION REGRESSION MODEL

EXPLANATORY VARIABLES	COEFF	t	PROB VALUE
PERCENTAGE MANUFACTURING INDUSTRY ..	-30.7	0.2	.817
AVERAGE FAMILY SIZE	-62.4	-1.4	.172
PRIMARY MALE MIL AVAILABLE 17-21/10	5.8	8.3	.001
UNEMPLOYMENT-BLACK MALE	-4.1	-1.8	.081
POPULATION CHANGE 1970 - 1980	3.2	0.1	.959
MEAN WAGES	5.6	0.7	.491
MILITARY INSTALLATIONS COUNT	-2.0	-1.7	.098
INTERCEPT TERM	238.0	-1.8	.081

Dependent variable-Nonprior Service Accession Counts

N = 48

R SQUARE = .741

TABLE 12

MID-LEVEL ACCESSION REGRESSION MODEL

EXPLANATORY VARIABLES	COEFF	t	PROB VALUE
PERCENTAGE MANUFACTURING INDUSTRY ..	-22.6	-0.9	.372
AVERAGE FAMILY SIZE	56.4	3.5	.001
PRIMARY MALE MIL AVAILABLE 17-21/10	5.1	41.6	.001
UNEMPLOYMENT-BLACK MALE	-2.6	-3.7	.001
POPULATION CHANGE 1970 - 1980	-142.5	-8.6	.001
MEAN WAGES	-3.1	-1.8	.069
MILITARY INSTALLATIONS COUNT	-1.5	-3.9	.001
INTERCEPT TERM	-118.5	-2.2	.029

Dependent variable-Nonprior Service Accession Counts
N = 742
R SQUARE = .805

3. **Hierarchical**

In this application of cluster analysis, the process yields a hierarchy of clus solutions, ranging from one overall cluster to as many clusters as there are cases. For this reason the file had to be reduced. Clusters at a higher level can contain several lower-level clusters, but within each level, the clusters are disjoint (each item belongs to only one cluster).

This example was constructed from a random sample of 40 Reserve Center local market areas out of 967. The same measurable factors (variables) used in the previous section are used to cluster the markets. Results should be similar but not the same.

Table 13 shows a list of results for the 3 cluster solution along with the actual accessions for each case. In the 3 cluster solution, cluster membership is as follows:

cluster 1: 29 market areas 158 average accessions

cluster 2: 8 market areas 1095 average accessions

cluster 3: 3 market areas 130 average accessions

Again these results suggest that local market areas according to accession counts can be classified as:

* high accession market areas (cluster 2);
* medium accession market areas (cluster 3) and;
* medium to low market areas (cluster 1).

TABLE 13

HIERARCHICAL CLUSTER RESULTS

CLUSTER MEMBERSHIP OF CASES USING AVERAGE LINKAGE (BETWEEN GROUPS)

ZIP	CASE	CLUSTERS '3'	ACCESSIONS
02905	1	1	523
11234	2	2	1892
14850	3	1	188
15401	4	1	317
16602	5	1	164
16652	6	1	140
18702	7	1	233
19007	8	2	1200
19013	9	2	1232
19090	10	2	1251
19401	11	2	1344
22701	12	1	29
24354	13	1	58
27101	14	1	81
28307	15	1	74
28712	16	1	97
32347	17	1	2
32570	18	1	77
32601	19	1	49
32812	20	1	110
36803	21	1	108
37662	22	1	100
40356	23	1	119
40505	24	1	125
43326	25	1	55
45431	26	1	122
55107	27	1	730
61614	28	1	189
62837	29	1	20
74074	30	1	37
75149	31	1	457
77701	32	1	74
83440	33	3	40
84062	34	3	288
88001	35	3	63
91105	36	2	884
92410	37	1	190
94965	38	2	592
98199	39	2	371
99207	40	1	134

Another output of hierarchical cluster analysis is the dendrogram. The key to reading a dendrogram is the concept of cluster level. By specifying a cluster level, the following information can be read from a dendrogram: the number of clusters and the Reserve Center markets contained in each cluster. That is, there is a correspondence

from cluster level to a partition of the Reserve Centers. Figure 3.1 is a display in graphic format (dendrogram) of the 40 market areas that were involved in the hierarchical clustering.

```
            DENDROGRAM USING AVERAGE LINKAGE (BETWEEN GROUPS)
                              RESCALED DISTANCE CLUSTER COMBINE

  C A S E    0         5        10        15        20        25
LABEL  SEQ   +---------+---------+---------+---------+---------+

19090   10    -+
19401   11    -+
19013    9    -+---------+
19007    8    -+         +-----+
91105   36    -----------+     +---------------+
94965   38    ---------+-------+               +-----+
98199   39    ---------+                       I     I
11234    2    ---------------------------------+     I
14850    3    ---+-+                                 I
61614   28    ---+ +-+                               I
16602    5    -+-+ I I                               I
16652    6    -+ +-+ I                               I
18702    7    -+-+   +-------+                       I
43326   25    -+     I       I                       I
15401    4    -----+ I       I                       I
45431   26    -----+-+       I                       +---------+
02905    1    -----+         +---+                   I         I
37662   22    -+-+           I   I                   I         I
77701   32    -+ +-+         I   I                   I         I
55107   27    ---+ +-------+ I   +-----+             I         I
75149   31    -----+       +-+   I     I             I         I
27101   14    -----+-------+     I     I             I         I
28712   16    -----+             I     I             I         I
62837   29    -------------------+     I             I         I
74074   30    ---+---+                 I             I         I
99207   40    ---+   +-------+         +-------------+         I
32812   20    ---+---+       I         I                       I
92410   37    ---+           I         I                       I
24354   13    -+-+           +---+     I                       I
40356   23    -+ +---+       I   I     I                       I
28307   15    ---+   +---+   I   I     I                       I
32347   17    ---+-+ I   I   I   I     I                       I
40505   24    ---+ +-+   +---+   +-----+                       I
22701   12    ---+-+     I       I                             I
32570   18    ---+       I       I                             I
36803   21    -----------+       I                             I
32601   19    -------------------+                             I
83440   33    -------+-----------+                             I
84062   34    -------+           +-----------------------------+
88001   35    -------------------+
```

Figure 3.1 Accession Dendrogram.

The scale at the top of the dendrogram is a cluster level scale. Note that the minimum value of cluster level is 25 at the far right. This corresponds with the highest

similarity coefficient (see Appendix I). A low cluster level specifies a partition having many small clusters while a high cluster level specifies a partition having a few large clusters. Thus, cluster level can be thought of as a measure of the largest dissimilarity (or, equivalently the weakest bond) present within any cluster in the partition.

An extremely useful part of the output is the visibility of data on markets. For example, consider cluster level 0, the minimum observed cluster level in the dendrogram. At cluster level 0, the 40 Reserve Center markets are partitioned into 32 clusters. Twenty-eight of these clusters contain only a single Reserve Center. Two of the 32 contain exactly 2 Reserve Centers and two clusters contain 4 Centers; identified by zipcodes 19090, 19401, 19013, 19007, and 16602, 16652, 18702, 43326. Since 0 is the minimum observed cluster level, we may conclude that the strongest possible bonds exist within every cluster. Specifically, we may conclude that the Reserve Center markets mentioned above are bound together by the tightest possible market ties. The cluster analysis will not separate them even at the lowest cluster level.

Consider next a slightly higher cluster level, say 11 Here we are permitting slightly weaker bonds to be present within clusters. We find that the 40 Reserve Centers are now partitioned into 5 clusters. One of these clusters contain a single Reserve Center (11234) and one contains 3 Reserve Centers. In the 7 Reserve Center cluster 19090, 19401, 19013, and 19007 have been joined by 91105, 94165, and 98199. It may be concluded that slightly weaker ties bind the new Reserve Centers to the original five. Similar inferences can can be drawn from other dendrograms, using different measurable variables.

It is not until level 17 that a 3 cluster solution is apparent. The dendrogram provides visibility to the broad scope of bonds that bring Reserve Center markets together. From Table 14 one could deduce that it is the high mean values of primary military available and military installations counts, along with low population change which causes markets 19090, 19401, 19013, and 19007 to join initially and form cluster 2. Market areas 83440, 84062, and 88001 cluster at level 10 and do not allow others to join until the final one cluster solution. This may be cause for further investigation of these markets.

C. FILL RATE

1. Nonhierarchical

This section analyzes fill rate, defined as the number assigned divided by the number authorized (assigned/authorized) for each Reserve Center market. Fill rates are

TABLE 14

HIERARCHICAL CLUSTER CHARACTERISTICS

	CLUSTER 1 N = 29		CLUSTER 2 N = 8		CLUSTER 3 N = 3	
VARIABLE	MEAN	STANDARD DEVIATION	MEAN	STANDARD DEVIATION	MEAN	STANDARD DEVIATION
MANUFACT	.3	.07	.3	.03	.2	.03
AVGFAMSZ	3.3	.09	3.3	.11	3.7	.06
PMILAVAL	31.3	25.90	249.4	148.00	18.5	17.71
UNEMPB	5.0	2.99	6.7	1.00	4.3	4.51
POPCHNGE	0.2	.13	.0	.07	.4	.09
WAGES	7.8	1.84	9.2	.82	6.8	.98
MILINS	3.7	5.31	45.5	13.41	1.3	1.16

important indicators of unit readiness levels. To demonstrate the application of cluster analysis for fill rate, a three cluster analysis for the total sample (967) was run. This is a departure from the previous method because the lone measurable factor influencing the cluster result is fill rate. Table 15 shows the results of this 3 cluster analysis.

TABLE 15

FILL RATE CLUSTER RESULTS

FINAL CLUSTER CENTERS.

CLUSTER	ZFILLRAT	NUMBER CASES	FILLR MEAN	STD DEV
1	4.171	8	1.855	.360
2	-1.177	247	.696	.157
3	.363	709	1.030	.126

The mean fill rates in the three cluster solution suggests the expected fill rate classification of:

* 8 high fill rate markets (cluster 1);
* 709 medium fill rate markets (cluster 3) and;
* 247 low fill rate markets (cluster 2).

2. Regression

To develop a predicting equation for fill rates variables in the data file were used to estimate a regression model for explaining variations in Reserve Center fill rates. Final variables found to be statistically significant and used in the model are:

unemployment-black males;

population change;

authorized billets and;

number of USA (active) recruiters.

The regression model is shown in Table 16.

TABLE 16

FILL RATE REGRESSION MODEL

EXPLANATORY VARIABLES	COEFF	t	PROB VALUE
NUMBER OF USA (ACTIVE) RECRUITERS ..	.16	4.3	.001
UNEMPLOYMENT-BLACK MALE	.05	1.6	.107
POPULATION CHANGE 1970 - 1980	.14	4.0	.001
AUTHORIZED BILLETS..................	-.16	-4.5	.001
INTERCEPT TERM	3.46	.1	.991

Dependent Variable=FILL RATE

N = 967

R SQUARE = .4

All of the variables used are significant at the 1% level except for black male-unemployment, which is significant at the 10% level. Results of this model suggests that high fill-rates can be explained by relatively high unemployment black males, a relatively high number of active recruiters, and positive population growth. A relatively low number of authorizations would be associated with high fill-rates.

A separate regression of fill rates for medium and low fill rate market areas was calculated to see if the nonhierarchical clusters using only fill rate as a measurable variable could be explained by the independent regression variables. The high fill rate market was not analyzed due to a small sample (8) size. The results are shown in Tables 17 and 18.

TABLE 17

LOW FILL RATE (CLUSTER 2) REGRESSION MODEL

EXPLANATORY VARIABLES	COEFF	t	PROB VALUE
AUTHORIZED BILLETS.................	.08	1.5	.126
POPULATION CHANGE 1970 - 1980	.02	-.4	.681
UNEMPLOYMENT-BLACK MALE	-.08	-1.7	.100
NUMBER OF USA (ACTIVE) RECRUITERS ..	-.08	-1.1	.275
INTERCEPT TERM	.63	-23.9	.001

Dependent Variable=FILL RATE

N = 247

R SQUARE = .2

TABLE 18

MEDIUM FILL RATE (CLUSTER 3) REGRESSION MODEL

EXPLANATORY VARIABLES	COEFF	t	PROB VALUE
AUTHORIZED BILLETS.................	-.13	-5.5	.001
POPULATION CHANGE 1970 - 1980	.09	4.5	.001
UNEMPLOYMENT-BLACK MALE	.01	.5	.652
NUMBER OF USA (ACTIVE) RECRUITERS ..	.06	2.6	.008
INTERCEPT TERM	.35	16.7	.001

Dependent Variable=FILL RATE

N = 709

R SQUARE = .61

These results are not encouraging as a predicting model for market fill rates. Except for black male unemployment, which is significant at the 10% level, variables in the low fill rate (cluster 2) regression model are not individually statistically significant in explaining low fill rate markets. The cluster 3 regression model shows some promise of being a good predicter, because authorized billets, population change, and number of USA active recruiters are significant at the 1% level. Unemployment-black males is not statistically significant for cluster 3.

A nonhierarchical cluster analysis using the four measurable variables for predicting all fill rates yields the results shown in Table 19. The results have a much closer range between means for low markets and medium fill rate markets (.02), which would suggest there is not much difference between the two clusters. Also the high fill rate cluster (cluster 3) has a large number of market areas (N = 91) compared to the previous straight fill-rate cluster where N = 8. The disparity is likely caused by the lack of significance in explaining individual clusters.

TABLE 19

NONHIERARCHICAL CLUSTER CHARACTERISTICS

	CLUSTER 1 N = 867 mean fill=.94		CLUSTER 2 N = 9 mean fill=.92		CLUSTER 3 N = 91 mean fill=1.1	
PREDICT VARIABLE	MEAN	STANDARD DEVIATION	MEAN	STANDARD DEVIATION	MEAN	STANDARD DEVIATION
ACTREC	205.1	142.31	16.9	12.20	23.9	22.85
UNEMPB	6.6	1.59	7.2	5.12	4.0	3.06
POPCHNGE	0.0	.06	.3	.17	.2	.14
AUTHORIZ	9.3	1.32	7.0	1.07	7.7	1.32
MILINS	40.2	22.39	4.6	6.35	4.0	6.63

3. **Hierarchical**

To better understand the interactions of fill rate clusters, a hierarchical cluster using the reduced sample of 40 markets and fill rate as the lone measurable variable is presented. These results are in Table 20.

Table 20 sheds some light on why there is a lack of explanation for fill rate clusters. Cluster 2 contains only one market area, identified by zipcode 32347, and cluster 3 contains only two market areas, identified by zip codes 94965 and 43326. Although the mean fill rates for the three clusters can be classified as:

high (cluster 2 mean = 2.04),
medium (cluster 1 mean = .97),
low (cluster 3 mean = .42).

A large portion (37) of the fill rate markets fall into the same cluster, indicating there is no real distinguishable dissimilarity between fill rates in the markets. A clearer picture can be viewed with the help of the dendrogram shown in Figure 3.2.

TABLE 20

HIERARCHICAL FILL RATE CLUSTERS

CLUSTER MEMBERSHIP OF CASES USING
AVERAGE LINKAGE (BETWEEN GROUPS)

NUMBER OF CLUSTERS=3

LABEL	CASE	CLUSTER	FILLRATE
02905	1	1	.82
11234	2	1	1.09
14850	3	1	1.20
15401	4	1	.91
16602	5	1	1.07
16652	6	1	.78
18702	7	1	.84
19007	8	1	1.00
19013	9	1	1.00
19090	10	1	.91
19401	11	1	.90
22701	12	1	1.02
24354	13	1	.87
27101	14	1	.81
28307	15	1	1.00
28712	16	1	.91
32347	17	2	2.04
32570	18	1	1.16
32601	19	1	.95
32812	20	1	1.01
36803	21	1	1.24
37662	22	1	1.05
40356	23	1	1.00
40505	24	1	1.00
43326	25	3	.58
45431	26	1	1.02
55107	27	1	.97
61614	28	1	.92
62837	29	1	1.00
74074	30	1	1.00
75149	31	1	.81
77701	32	1	.90
83440	33	1	1.03
84062	34	1	.99
88001	35	1	.92
91105	36	1	1.01
92410	37	1	.98
94965	38	3	.26
98199	39	1	.89
99207	40	1	.90

The dendrogram shows that on the lowest level, the 37 market areas immediately cluster together with market areas 43326, 94965, and 32347 staying alone as single market clusters. It is not until the final level, when the clustering algorithm forms a single cluster, that market area 32347 joins another market. This is a clear indication that this particular market is an outlier and should be disregarded from analysis.

```
                  DENDROGRAM USING AVERAGE LINKAGE (BETWEEN GROUPS)
                                           RESCALED DISTANCE CLUSTER COMBINE
   C A S E      0         5        10        15        20        25
  LABEL  SEQ    +---------+---------+---------+---------+---------+

  62837   29    -+
  74074   30    -+
  19007    8    -+
  28307   15    -+
  40356   23    -+
  19013    9    -+
  40505   24    -+
  32812   20    -+
  91105   36    -+
  55107   27    -+
  92410   37    -+
  84062   34    -+
  22701   12    -+
  45431   26    -+
  83440   33    -+
  37662   22    -+
  11234    2    -+
  16602    5    -+
  27101   14    -+
  75149   31    -+-+
  02905    1    -+ I
  18702    7    -+ I
  16652    6    -+ I
  77701   32    -+ I
  99207   40    -+ I
  19401   11    -+ I
  98199   39    -+ I
  61614   28    -+ +---------+
  88001   35    -+ I         I
  15401    4    -+ I         I
  28712   16    -+ I         I
  19090   10    -+ I         I
  24354   13    -+ I         +-----------------------------------+
  32601   19    -+ I         I                                   I
  14850    3    -+ I         I                                   I
  36803   21    -+-+         I                                   I
  32570   18    -+           I                                   I
  43326   25    -----+-------+                                   I
  94965   38    -----+                                           I
  32347   17    -------------------------------------------------+
```

Figure 3.2 Fill Rate Dendrogram.

D. COMPETITIVE SUCCESS

1. Nonhierarchical

To demonstrate this application of cluster analysis for competitive success, a three cluster analysis for the total sample (967) was run. Competitive success is defined

as the number of Army Reserve accessions divided by the number of National Guard Accessions. The definition of competitive success naturally underscores the fact that successful Army Reserve markets will have high values for the success variable. If the three cluster solution clusters based on high low and medium success, one could infer a classification of clusters based on relative success. Results of the clustering are shown in Table 21.

TABLE 21

COMPETITIVE CLUSTER RESULTS

FINAL CLUSTER CENTERS.

CLUSTER	MEAN SUCCESS	NUMBER CASES	STD DEV
1	2.965	852	1.451
2	8.965	106	3.431
3	39.750	1	-
	MISSING	8	

The mean success values in the three cluster solution suggests the expected success classifications of high, low, and medium, but, cluster 3 contains only one market area out of 959 (8 cases are missing). It is apparent that cluster 3 contains an outlier market which will not be further considered in this analysis. This leaves two clusters to consider for classification (cluster 1 and cluster 2). The means of the two clusters lends itself to the following classification:

* high competitive success markets (cluster 2)
* moderate competitive success markets (cluster 1)

2. **Regression**

Predicting equations for high competitive success markets and for moderate competitive success markets were developed Again, variables from the data file were used to find a significant model for each of the classifications. Table 22 shows the regression model results for predicting high competitive success markets. The following variables were used because of their statistical significance:

percentage manufacturing industry;

median rent;

average family size; and

military installations count.

TABLE 22

HIGH COMPETITIVE SUCCESS MODEL

EXPLANATORY VARIABLES	COEFF	t	PROB VALUE
PERCENTAGE MANUFACTURING INDUSTRY ..	-13.14	-3.4	.001
MEDIAN RENT	-.02	-2.0	.044
AVERAGE FAMILY SIZE	-3.19	-1.6	.100
MILITARY INSTALLATIONS COUNT	-.06	-2.2	.030
INTERCEPT TERM	28.83	3.9	.001

Dependent Variable=Relative Success

N = 106

R SQUARE = .3

The variables median rent and military installations count are statistically significant at the 5% level. While percentage manufacturing industry is significant at the 1% level and average family size is significant at the 10% level. The coefficients of each of the variable are negative which suggests high success rates correspond with or can be explained by low or negative values for each variable.

Table 23 shows the variables used to develop a moderate competitive success model:

percentage government industry;

mean wages;

population change;

number of USAR recruiters.

As shown in Table 23, all of the variables used to explain moderate competitive success are statistically significant at the 1% level. These markets, based on the model, are likely to be located where the percentage of government industry is low, mean wages are high, population change is negative or low, and the number of Army Reserve recruiters is high.

TABLE 23

MODERATE COMPETITIVE SUCCESS MODEL

EXPLANATORY VARIABLES	COEFF	t	PROB VALUE
PERCENTAGE GOVERNMENT INDUSTRY	-7.13	-3.9	.001
MEAN WAGES	.09	2.7	.001
POPULATION CHANGE 1970 - 1980	-1.01	-2.7	.001
NUMBER OF USAR RECRUITERS	.03	7.9	.001
INTERCEPT TERM	2.37	8.0	.001

Dependent Variable=Moderate Competitive Success

N = 852

R SQUARE = .28

3. **Hierarchical**

Table 24 shows a hierarchical cluster result using competitive success as the lone measurable variable. The file was scaled down from 967 markets to a random sample of 40 Reserve Center markets.

Cluster 3 contains only two market areas, identified by zip codes 77701, and 61614. Cluster 2 contains nine market areas and cluster 1 has 29 market areas. It is clear that cluster 1 dominates this three cluster algorithm. Cluster 3, although associated with high success values appear to be outliers. The mean success value for cluster 3 is 6.36, the mean for cluster 2 is 5.9, and the mean for cluster 1 is 2.75. If cluster 3 is an outlier, these results correspond with the results obtained earlier using the nonhierarchical method. Cluster 2 would be classified as high success markets and cluster 1 would be classified as moderate success markets. Cluster 3 would be dropped from analysis. A clearer picture can be viewed with the help of the dendrogram shown in Figure 3.3.

The dendrogram shows that on the lowest level, the 29 market areas of cluster 1 immediately cluster together. Market areas 61614 and 77701 join at level two and remain away from the others until the final level forms a single cluster. This is further evidence that these two markets are outliers.

TABLE 24

HIERARCHICAL COMPETITIVE SUCCESS CLUSTERS

CLUSTER MEMBERSHIP OF CASES USING
AVERAGE LINKAGE (BETWEEN GROUPS)

CLUSTERS=3

LABEL	CASE	3	COMPETITIVE SUCCESS
02905	1	1	1.23
11234	2	1	2.28
14850	3	2	3.66
15401	4	2	2.45
16602	5	1	1.70
16652	6	1	1.57
18702	7	1	1.96
19007	8	2	3.01
19013	9	2	2.81
19090	10	2	3.00
19401	11	2	2.76
22701	12	1	1.06
24354	13	1	2.49
27101	14	1	1.52
28307	15	1	1.31
28712	16	1	1.87
32347	17	1	.62
32570	18	1	1.29
32601	19	1	2.72
32812	20	1	3.05
36803	21	1	1.03
37662	22	1	.73
40356	23	1	1.40
40505	24	1	1.43
43326	25	1	1.19
45431	26	1	1.94
55107	27	2	3.12
61614	28	3	4.34
62837	29	1	.63
74074	30	1	1.07
75149	31	2	4.20
77701	32	3	8.38
83440	33	1	1.36
84062	34	1	.97
88001	35	1	1.35
91105	36	1	1.61
92410	37	1	1.07
94965	38	2	4.75
98199	39	1	1.44
99207	40	1	2.51

E. OVERVIEW

The application of a particular clustering scheme to a particular set of data involves assumptions about the appropriateness of the statistical and mathematical techniques employed in the scheme. These assumptions are often difficult to justify and the researcher must rely to some extent on intuition and experience with the

```
             DENDROGRAM USING AVERAGE LINKAGE (BETWEEN GROUPS)
                              RESCALED DISTANCE CLUSTER COMBINE
 C A S E    0         5        10        15        20        25
LABEL  SEQ  +---------+---------+---------+---------+---------+

19090   10   -+
55107   27   -+
19007    8   -+
75149   31   -+
94965   38   -+
19013    9   -+
19401   11   -+---+
15401    4   -+   I
14850    3   -+   I
16652    6   -+   I
45431   26   -+   I
11234    2   -+   I
16602    5   -+   +-------------------------------------------+
24354   13   -+   I                                           I
32812   20   -+   I                                           I
40356   23   -+   I                                           I
40505   24   -+   I                                           I
32601   19   -+   I                                           I
83440   33   -+   I                                           I
18702    7   -+---+                                           I
28712   16   -+                                               I
99207   40   -+                                               I
32347   17   -+                                               I
62837   29   -+                                               I
36803   21   -+                                               I
74074   30   -+                                               I
27101   14   -+                                               I
43326   25   -+                                               I
02905    1   -+                                               I
84062   34   -+                                               I
91105   36   -+                                               I
37662   22   -+                                               I
88001   35   -+                                               I
92410   37   -+                                               I
22701   12   -+                                               I
98199   39   -+                                               I
28307   15   -+                                               I
32570   18   -+                                               I
61614   28   -----+-------------------------------------------+
77701   32   -----+
```

Figure 3.3 Competitive Success Dendrogram.

characteristics of the objects under consideration. It would be unwise to accept these results uncritically It is possible that other approaches to cluster analysis would be more appropriate or yield better results. However, each clustering scheme produced a cluster of Reserve Center markets related to success, such as, high accessions, high fill rates, and high competitive success. Also, the different schemes between markets do not yield a consistent or discernable pattern. High accession clusters are not the same markets as high fill rates and or high competitive success.

Disappointing results from the regression models for fill rates does not mean that cluster analysis is the wrong approach; indeed, it may mean there were no predicting variables in the data set or that fill rates are not predictable dependent variables, or that other notions of similarity should be explored.

Accession counts were chosen for analysis first because of the abundance of literature and proven theories on the subject. The cluster results according to accession counts can assist Army Reserve recruiters in identifying markets where accessions would be expected to be high or low. Low accessions in medium market clusters may serve as criteria for future locations of Reserve units. Fill rates measure how successful the markets are in reaching their goals (authorizations). Cluster results may indicate market areas (low) where more recruiting resources should be increased or market areas (high) where recruiting resources can be relaxed. Competitive success clusters can assist policy makers on decisions to expand, where to locate new units, and allocation of recruiting resources. Moderate competitive success markets indicate that more recruits are enlisting in the National Guard rather than the Army Reserves. The success of National Guard recruiting may indicate fertile ground for expanding current units or adding additional units.

Overall, cluster analysis applied to accession modeling is very encouraging. Policy makers can classify and identify Reserve Center markets according to accession counts, fill rates, and competitive success. The classifications could play an important role in the location of future Reserve units or in the expansion of current units. In view of the encouraging results that have been achieved using this data base, a next step would be to expand the base to provide new and different avenues for analysis. One avenue to pursue would be to include data in the accession data base which will be responsive to changes in local areas or variables which reflect the military propensity of local markets. Willingness to serve can play an important role in the success or failure in obtaining future accessions.

IV. CONCLUSION

This thesis has discussed the use of cluster analysis to group Reserve Centers based on the economic characteristics of Reserve Center local labor markets and characteristics of individual Reserve Centers. The analysis has presented three examples of cluster analysis in which Reserve Center market characteristics are treated as measurable objects. The results obtained in each demonstration are not presented as solid conclusions; they are implied incidentally while demonstrating applications of clustering methods to manpower problems. It is asserted that the methods used here are representative of a wide range of applications for cluster analysis in the area of manpower planning and Reserve Center classification. There is no "correct" way to cluster data and a variety of methods are available, each requiring a different set of assumptions and utilizing different aspects of the measurements as the basis for discrimination between groups.

Although cluster analysis was developed for the physical sciences, it can have a wide range of applications in manpower analysis. Potential users should be aware of three concluding precautionary generalizations outlined by Aldenderfer: [Ref. 14: p. 15]:

(1) The strategy of cluster analysis is structure-seeking although its operation is structure-imposing.

That is, clustering methods are used to discover structure in data that is not readily apparent by visual inspection or other methods. Although the strategy of clustering may be structure seeking, its operation is one that is structure imposing. A clustering method will always place objects into groups, and these groups may be radically different in composition when different clustering methods are used. The key to using cluster analysis is knowing when these groups are real and not merely imposed on the data by the method. A number of validation procedures have been developed to provide some relief for this problem.

(2) Cluster analysis methods have evolved from many disciplines and are inbred with the biases of these disciplines.

Each discipline has its own biases and preferences as to the kinds of questions asked of the data, the types of data thought to be useful in building a classification, and the

structure of classifications thought to be useful. Since clustering methods are often no more than plausible rules for creating groups, manpower users must be aware of the biases that often accompany the presentation and description of a clustering method.

(3) Cluster analysis methods are relatively simple procedures that in many cases, are not supported by an extensive body of statistical reasoning.

In other words, many cluster analysis methods are heuristics (simple rules of thumb). They are little more than plausible algorithms that can be used to create clusters of cases. This stands in sharp contrast to factor analysis, for instance, which is based upon an extensive body of statistical reasoning. Although many clustering algorithms have important mathematical properties that have been explored in some detail, it is important to recognize the fundamental simplicity of many of the methods. In doing so, the user is far less likely to make the mistake of reifying the cluster solution.

The three applications of cluster analysis used in this thesis suggests that:

(1) Reserve Center Markets according to accession counts can be classified and identified as: high accession market areas, medium accession market areas, and low accession market areas.

(2) Reserve Center Markets according to fill rate can be classified as: high fill rate markets, medium fill rate markets, and low fill rate markets.

(3) Reserve Center markets according to competitive success measures can be classified as: high competitive success markets, and moderate competitive success markets.

Accession count clusters, fill rate clusters, and competitive success clusters can assist policy formulation in the areas of: location of new units, allocation of new authorizations, and allocation of recruiting resources. In addition, cluster analysis can serve as a source for hypotheses about accessions which can be tested using regression and other multivariate analysis.

APPENDIX A
VARIABLE DEFINITION

NPS Male/Female... Cumulative counts of all NPS accessions in the appropriate category during FY 83-85 inclusive. NPS implies no prior service in any Active or Reserve component.

QMA Male/Female... Qualified military available. This is a count of the male female population in the market area aged 17-29 years.

% of Population Black... Total number of blacks divided by the total population in each market in 1980. (all ages and sexes).

Average Family Income... Average income accruing to all families in the market area from all sources in 1980.

Average Family Size... Average number of family members in 1980.

Median Home Value... Median value of all family homes in the market area in 1980.

Median Home Rent... Median rent paid for all dwellings in the market area in 1980.

% of Families with Dual Workers... Number of families with two or more members holding full or part time jobs in 1980.

% Population Change... Total population figures for each market area. ((1980-1970) 1970)x100.

Manufacturing Workers... Proportion of workers reported in census classifications 'manufacturing', 'transport' and 'communications' in 1980 in each market area.

Service Workers... Proportion of workers reported in census classifications 'wholesale', 'retail', 'finance', 'service', 'recreation', 'health', 'education', and 'other' in 1980 in each market area.

Government Workers... Proportion of workers reported in 'government' census classification in 1980 in each market area.

Seasonal Workers... Proportion of workers reported in census classifications 'agriculture' and 'construction' in 1980 in each market area.

APPENDIX B

VARIABLE LISTING

ZIPCODE 'CENTROID ZIPCODE FOR MARKET AREA'
POPCHNGE 'POPULATION CHANGE 1970 - 1980'
BLUE 'PERCENTAGE BLUE COLLAR OCCUPATION'
WHITE 'PERCENTAGE WHITE COLLAR OCCUPATION'
FARMING 'PERCENTAGE FARMING OCCUPATION'
MANUFACT 'PERCENTAGE MANUFACTURING INDUSTRY'
SEASONAL 'PERCENTAGE SEASONAL INDUSTRY'
GOVERN 'PERCENTAGE GOVERNMENT INDUSTRY'
SERVICE 'PERCENTAGE SERVICE INDUSTRY'
BLACK 'BLACK POPULATION PERCENTAGE'
UNEMPW 'UNEMPLOYMENT-WHITE MALE'
UNEMPB 'UNEMPLOYMENT-BLACK MALE'
PMILAVAL 'PRIMARY MALE MILITARY AVAILABLE 17-21 1000'
SMILAVAL 'SECONDARY MALE MILITARY AVAILABLE 22-29'
TMILAVAL 'TOTAL MALE MILITARY AVAILABLE'
TMFAVAL 'TOTAL MALE & FEMALE MILITARY AVAILABLE'
AVGFAMIN 'AVERAGE FAMILY INCOME'
AVGFAMSZ 'AVERAGE FAMILY SIZE'
TWOWRK 'FAMILIES WITH 2 OR MORE WORKERS'
HOMEVAL 'MEDIAN HOME VALUE'
RENT 'MEDIAN RENT'
RECRUIT 'NUMBER OF USAR RECRUITERS'
ACTREC 'NUMBER OF USA (ACTIVE) RECRUITERS'
LOSSES 'NUMBER ATTRITED'
MILINS 'MILITARY INSTALLATIONS COUNT'
PSQMA 'PRIOR SERVICE MALE COUNT'
AUTHORIZ 'AUTHORIZED BILLETS'
ASSIGNED 'ASSIGNED PERSONNEL'
USARMEMB 'USAR MEMBERS'
GUARDMEM 'GUARD MEMBERS'
TCOMMEMB 'TOTAL RESERVE COMPONENTS MEMBER6'
ADAUTHOR 'ADJUSTED AUTHORIZATION'
OVERLAP 'OVERLAP FLAG'
BRIGADE 'BRIGADE'
BRIGADE2 'DUMMY VARIABLE FOR BRIGADE 2'
BRIGADE4 'DUMMY VARIABLE FOR BRIGADE 4'
BRIGADE5 'DUMMY VARIABLE FOR BRIGADE 5'
BRIGADE6 'DUMMY VARIABLE FOR BRIGADE 6'
MISSNPS '(AUTHORIZE - ASSIGNED + LOSSES) PMILAVAL'
RECRPMIL 'RECRUIT PRIMARY MILITARY AVAILABLE'
MNPSPMIL 'ARMY NPS MALES 17-21 PRIMARY MILITARY AVAILABLE'

```
ARECPMIL 'ACTIVE RECRUIT PRIMARY MILITARY AVAILABLE'
GUARPMIL 'GUARD MEMBERS PRIMARY MILITARY AVAILABLE'
MISSPS   '(AUTHORIZE - ASSIGNED + LOSSES) PSQMA'
RECRPQMA 'RECRUIT PRIOR SERVICE MIL AVAILABLE'
MPSPQMA1 'ARMY MALE PS 17-21 PRIOR SERVICE MIL AVAIL'
MPSPQMA2 'ARMY MALE PS ALL AGES PRIOR SERVICE MIL AVAIL'
ARECPQMA 'ACTIVE RECRUITERS PRIOR SERVICE MIL AVAIL'
GUARPQMA 'GUARD MEMBERS PRIOR SERVICE MIL AVAILABLE'
GAINS    '= JOINED RESERVES '
ARECRS   '= ACTIVE RECRUITERS ALL BRANCHES'
UNEMP    'UNEMPLOYMENT RATE FOR THIS MARKET'
WAGES    'MEAN WAGES'
MISSNPST '(AUTHORIZ - ASSIGNED + LOSSES) TMILAVAL'
RECRTMIL 'RECRUIT TMILAVAL'
MNPSTMIL 'ANPSMP TMILAVAL'
MNPATMIL 'ANPSMA TMILAVAL'
ARECTMIL 'ACTREC TMILAVAL'
GUARTMIL 'GUARDMEM TMILAVAL'
ADJ.AUTH 'ADAUTHOR AUTHORIZ'
```

APPENDIX C
ACCESSION VARIABLES

ANPSAP 'ARMY NONPRIOR SERVICE MALE & FEMALE 17-21'
ANPSAS 'ARMY NONPRIOR SERVICE MALE & FEMALE 22-29'
ANPSAA 'ARMY NONPRIOR SERVICE MALE & FEMALE ALL AGES'
ANPSMP 'ARMY NONPRIOR SERVICE MALE 17-21'
ANPSMS 'ARMY NONPRIOR SERVICE MALE 22-29'
ANPSMA 'ARMY NONPRIOR SERVICE MALE ALL AGES'
APSAP 'ARMY PRIOR SERVICE MALE & FEMALE 17-21'
APSAS 'ARMY PRIOR SERVICE MALE & FEMALE 22-29'
APSAA 'ARMY PRIOR SERVICE MALE & FEMALE ALL AGES'
APSMP 'ARMY PRIOR SERVICE MALE 17-21'
APSMS 'ARMY PRIOR SERVICE MALE 22-29'
APSMA 'ARMY PRIOR SERVICE MALE ALL AGES'
GNPSAP 'GUARD NONPRIOR SERVICE MALE & FEMALE 17-21'
GNPSAS 'GUARD NONPRIOR SERVICE MALE & FEMALE 22-29'
GNPSAA 'GUARD NONPRIOR SERVICE MALE & FEMALE ALL AGES'
GNPSMP 'GUARD NONPRIOR SERVICE MALE 17-21'
GNPSMS 'GUARD NONPRIOR SERVICE MALE 22-29'
GNPSMA 'GUARD NONPRIOR SERVICE MALE ALL AGES'
GPSAP 'GUARD PRIOR SERVICE MALE & FEMALE 17-21'
GPSAS 'GUARD PRIOR SERVICE MALE & FEMALE 22-29'
GPSAA 'GUARD PRIOR SERVICE MALE & FEMALE ALL AGES'
GPSMP 'GUARD PRIOR SERVICE MALE 17-21'
GPSMS 'GUARD PRIOR SERVICE MALE 22-29'
GPSMA 'GUARD PRIOR SERVICE MALE ALL AGES'
ONPSAP 'OTHER NONPRIOR SERVICE MALE & FEMALE 17-21'
ONPSAS 'OTHER NONPRIOR SERVICE MALE & FEMALE 22-29'
ONPSAA 'OTHER NONPRIOR SERVICE MALE & FEMALE ALL AGES'
ONPSMP 'OTHER NONPRIOR SERVICE MALE 17-21'
ONPSMS 'OTHER NONPRIOR SERVICE MALE 22-29'
ONPSMA 'OTHER NONPRIOR SERVICE MALE ALL AGES'
OPSAP 'OTHER PRIOR SERVICE MALE & FEMALE 17-21'
OPSAS 'OTHER PRIOR SERVICE MALE & FEMALE 22-29'
OPSAA 'OTHER PRIOR SERVICE MALE & FEMALE ALL AGES'
OPSMP 'OTHER PRIOR SERVICE MALE 17-21'
OPSMS 'OTHER PRIOR SERVICE MALE 22-29'
OPSMA 'OTHER PRIOR SERVICE MALE ALL AGES'
APSDAA 'ARMY PRIOR SERVICE DIRECT M & FEM ALL AGES'
APSDMA 'ARMY PRIOR SERVICE DIRECT MALE ALL AGES'

APPENDIX D
SAMPLE VALUE LISTINGS

ZIPCODE	ANPSAA	RECRUIT	PMILAVAL	MILINS
2905	563	24	92	25
11234	1892	73	531	73
14850	188	6	26	2
15401	317	10	44	0
16602	164	3	23	0
16652	140	3	20	0
18702	233	10	29	4
19007	1200	53	210	46
19013	1232	52	206	40
19090	1251	54	213	45
19401	1344	61	226	41
22701	29	2	18	5
24354	58	2	10	0
27101	81	2	36	2
28307	74	2	24	3
28712	97	5	26	0
32347	2	0	2	0
32570	77	2	19	13
32601	49	3	16	0
32812	110	5	39	7
36083	108	4	16	2
37662	100	2	22	0
40356	119	7	23	2
40505	125	7	23	2
43326	55	3	18	2
45431	216	6	54	8
74074	37	1	8	0
75149	457	21	94	9
77701	74	2	18	7
83440	40	1	5	0
84062	288	12	39	2
88001	63	3	12	2
91105	884	41	406	28
92410	190	6	79	8
94965	592	33	116	54
98199	371	16	87	37
99207	134	4	16	3

ZIPCODE	AUTHORIZ	ASSIGNED	BLUE	WHITE	BLACK
2905	360	295	.35	.63	.01
11234	871	948	.24	.75	.13
14850	109	131	.34	.57	.01
15401	363	330	.49	.49	.02
16602	231	247	.51	.45	.00
16652	147	115	.51	.44	.00
18702	282	237	.46	.52	.01
19007	694	694	.29	.69	.09
19013	243	242	.31	.67	.11
19090	424	387	.30	.68	.09
19401	149	134	.32	.66	.08
22701	80	82	.39	.52	.15
24354	223	195	.53	.43	.03
27101	570	464	.46	.50	.12
28307	407	407	.43	.51	.27
28712	65	59	.45	.51	.06
32347	27	55	.37	.51	.25
32570	69	80	.33	.63	.14
32601	399	378	.31	.62	.20
32812	337	340	.27	.65	.11
36083	95	118	.39	.56	.44
37662	304	318	.49	.48	.02
40356	28	28	.38	.52	.05
40505	232	231	.35	.55	.06
43326	121	70	.45	.46	.01
45431	140	143	.38	.57	.07
55107	120	117	.28	.68	.02
61614	523	483	.37	.52	.01
62837	30	30	.40	.46	.00
74074	141	141	.40	.52	.05
75149	449	365	.32	.66	.13
77701	359	322	.45	.52	.14
83440	96	99	.30	.49	.00
84062	69	68	.32	.65	.00
88001	158	146	.29	.62	.02
91105	175	176	.29	.70	.10
92410	289	283	.30	.67	.05
94965	229	60	.22	.77	.12
98199	1501	1337	.30	.68	.03
99207	381	342	.31	.62	.01

ZIPCODE	FARMING	MANUFACT	SEASONAL	GOVERN	SERVICE
2905	.02	.38	.07	.05	.51
11234	.01	.30	.04	.05	.61
14850	.08	.33	.13	.04	.49
15401	.02	.33	.19	.03	.45
16602	.04	.33	.22	.04	.42
16652	.06	.37	.18	.06	.40
18702	.03	.40	.10	.05	.45
19007	.02	.32	.07	.05	.56
19013	.02	.33	.08	.04	.55
19090	.02	.33	.08	.05	.55
19401	.02	.35	.08	.04	.53
22701	.08	.25	.22	.07	.46
24354	.05	.32	.26	.03	.38
27101	.04	.47	.10	.02	.41
28307	.06	.35	.14	.06	.45
28712	.04	.40	.12	.03	.44
32347	.12	.27	.22	.08	.43
32570	.03	.22	.12	.11	.55
32601	.07	.18	.18	.08	.56
32812	.07	.19	.18	.05	.58
36083	.04	.33	.11	.08	.49
37662	.04	.31	.24	.03	.42
40356	.10	.31	.18	.05	.46
40505	.10	.29	.18	.06	.47
43326	.08	.42	.14	.03	.41
45431	.05	.37	.10	.05	.47
55107	.03	.31	.09	.04	.56
61614	.11	.36	.18	.02	.44
62837	.14	.27	.27	.03	.43
74074	.07	.24	.25	.05	.46
75149	.02	.31	.12	.04	.54
77701	.03	.31	.20	.03	.45
83440	.21	.20	.30	.05	.46
84062	.03	.24	.13	.05	.57
88001	.09	.18	.19	.11	.52
91105	.01	.32	.06	.03	.58
92410	.03	.25	.12	.05	.58
94965	.02	.23	.08	.06	.64
98199	.02	.29	.10	.05	.56
99207	.08	.22	.17	.05	.57

ZIPCODE	AVGFAMIN	AVGFAMSZ	UNEMP	UNEMPB	UNEMPW
2905	21299	3	5	3	6
11234	25020	3	9	7	5
14850	17634	3	4	2	8
15401	18531	3	12	4	10
16602	16879	3	10	1	13
16652	16583	3	13	1	10
18702	17244	3	11	2	9
19007	24141	3	6	7	5
19013	23638	3	6	7	6
19090	24273	3	5	6	5
19401	23851	3	5	6	5
22701	18353	3	5	9	4
24354	14391	3	12	7	9
27101	18244	3	4	8	4
28307	14807	3	6	9	4
28712	16518	3	6	4	5
32347	13982	3	8	8	3
32570	15250	3	7	8	5
32601	14236	3	4	10	4
32812	17134	3	5	5	4
36083	14665	4	9	9	4
37662	14769	3	7	4	8
40356	14982	3	6	8	8
40505	15649	3	5	10	7
43326	19485	3	11	2	7
45431	20399	3	7	6	8
55107	24379	3	5	3	5
61614	21966	3	11	1	7
62837	16535	3	17	0	9
74074	16803	3	4	3	3
75149	22660	3	5	4	2
77701	20948	3	13	5	4
83440	15357	4	6	0	6
84062	19677	4	7	4	6
88001	15065	4	8	9	6
91105	23091	3	7	8	5
92410	20656	3	7	6	6
94965	24436	3	4	8	5
98199	22886	3	6	5	7
99207	19281	3	8	3	9

ZIPCODE	TWOWRK	HOMEVAL	RENT	POPCHNGE	WAGES
2905	.60	53707	254	.02	8
11234	.54	73340	313	-.1	8
14850	.56	36677	224	.01	8
15401	.42	37259	194	.03	7
16602	.43	31812	197	.03	7
16652	.48	32881	195	.07	7
18702	.50	38547	204	.05	7
19007	.56	60936	290	.0	8
19013	.55	57214	275	.0	9
19090	.56	61317	286	.0	9
19401	.56	59136	276	.0	9
22701	.58	52638	238	.35	7
24354	.45	32916	168	.16	6
27101	.62	42194	193	.15	9
28307	.57	35429	181	.20	6
28712	.55	41007	190	.17	10
32347	.48	25984	158	.18	8
32570	.48	34722	195	.19	7
32601	.47	37426	192	.54	7
32812	.51	51120	228	.46	8
36083	.51	33182	159	.22	6
37662	.42	33331	174	.17	9
40356	.53	38375	180	.20	7
40505	.56	40825	186	.20	8
43326	.56	42137	216	.07	8
45431	.54	47898	226	.00	8
55107	.64	69920	268	.07	10
61614	.54	47802	250	.09	9
62837	.48	30416	185	.08	7
74074	.52	32513	173	.29	6
75149	.62	56351	259	.29	11
77701	.49	35328	226	.09	10
83440	.58	46133	204	.31	7
84062	.57	67617	249	.40	8
88001	.49	46037	183	.48	6
91105	.55	106725	314	.10	10
92410	.50	78417	290	.34	8
94965	.54	116829	315	.03	9
98199	.53	74002	279	.16	10
99207	.51	55264	221	.25	8

APPENDIX E
Z-SCORE TRANSFORMATION

FROM LABEL	TO Z-SCORE	LABEL	VARIABLE
HOMEVAL	ZHOMEVAL	ZSCORE:	MEDIAN HOME VALUE
BLACK	ZBLACK	ZSCORE:	BLACK POPULATION PERCENTAGE
RECRUIT	ZRECRUIT	ZSCORE:	NUMBER OF USAR RECRUITERS
AUTHORIZ	ZAUTHORI	ZSCORE:	AUTHORIZED BILLETS
UNEMPB	ZUNEMPB	ZSCORE:	UNEMPLOYMENT-BLACK MALE
RENT	ZRENT	ZSCORE:	MEDIAN RENT
TWOWRK	ZTWOWRK	ZSCORE:	FAMILIES WITH 2 OR MORE WORKERS
AVGFAMSZ	ZAVGFAMS	ZSCORE:	AVERAGE FAMILY SIZE
TMILAVAL	ZTMILAVA	ZSCORE:	TOTAL MALE MILITARY AVAILABLE
UNEMPW	ZUNEMPW	ZSCORE:	UNEMPLOYMENT-WHITE MALE
GOVERN	ZGOVERN	ZSCORE:	PERCENTAGE GOVERNMENT INDUSTRY
ANPSAA	ZANPSAA	ZSCORE:	ARMY NONPRIOR SERVICE MALE & FE
SEASONAL	ZSEASONA	ZSCORE:	PERCENTAGE SEASONAL INDUSTRY
FILLRATE	ZFILLRAT	ZSCORE:	FILLRATE
MILINS	ZMILINS	ZSCORE:	MILITARY INSTALLATIONS COUNT
POPCHNGE	ZPOPCHNG	ZSCORE:	POPULATION CHANGE 1970 - 1980
PMILAVAL	ZPMILAVA	ZSCORE:	PRIMARY MALE MILITARY AVAILABLE
WAGES	ZWAGES	ZSCORE:	MEAN WAGES
MANUFACT	ZMANUFAC	ZSCORE:	PERCENTAGE MANUFACTURING INDUST
LOSSES	ZLOSSES	ZSCORE:	NUMBER ATTRITED
ACTREC	ZACTREC	ZSCORE:	NUMBER OF USA (ACTIVE) RECRUITE

APPENDIX F
Z-SCORE VALUES

ZIPCODE	ZHOMEVAL	ZBLACK	ZRECRUIT	ZAUTHORI
2905	.36563	-.66558	1.00665	.43133
11234	1.42539	.41145	4.45430	2.30626
14850	-.55363	-.66558	-.25983	-.48963
15401	-.52221	-.57583	.02161	.44234
16602	-.81623	-.75534	-.47091	.04199
16652	-.75853	-.75534	-.47091	-.35020
18702	-.45269	-.66558	.02161	.14513
19007	.75584	.05244	3.04710	1.65682
19013	.55493	.23195	2.97674	.00204
19090	.77640	.05244	3.11746	.66615
19401	.65868	-.03731	3.60998	-.34286
22701	.30792	.59096	-.54127	-.59603
24354	-.75664	-.48608	-.54127	-.07135
27101	-.25583	.32170	-.54127	1.20185
28307	-.62099	1.66799	-.54127	.60378
28712	-.31990	-.21682	-.33019	-.65107
32347	-1.13082	1.48849	-.68199	-.79050
32570	-.65916	.50121	-.54127	-.63639
32601	-.51320	1.03972	-.47091	.57442
32812	.22598	.23195	-.33019	.34694
36083	-.74228	3.19379	-.40055	-.54100
37662	-.73424	-.57583	-.54127	.22586
40356	-.46197	-.30657	-.18947	-.78683
40505	-.32973	-.21682	-.18947	-.03832
43326	-.25891	-.66558	-.47091	-.44560
45431	.05206	-.12706	-.25983	-.37588
55107	1.24078	-.57583	1.35845	-.44927
61614	.04688	-.66558	-.25983	1.02940
62837	-.89159	-.75534	-.61163	-.77949
74074	-.77839	-.30657	-.61163	-.37222
75149	.50835	.41145	.79557	.75788
77701	-.62645	.50121	-.54127	.42766
83440	-.04321	-.75534	-.61163	-.53733
84062	1.11647	-.75534	.16233	-.63639
88001	-.04839	-.57583	-.47091	-.30984
91105	3.22746	.14219	2.20277	-.24746
92410	1.69944	-.30657	-.25983	.17082
94965	3.77286	.32170	1.63989	-.04933
98199	1.46112	-.48608	.44377	4.61783
99207	.44967	-.66558	-.40055	.50838

ZIPCODE	ZRENT	ZTWOWRK	ZAVGFAMS	ZTMILAVA
2905	.73604	1.10985	.26255	.36437
11234	1.92327	.12222	.26255	5.28581
14850	.13237	.45143	.26255	-.33556
15401	-.47131	-1.85303	-.34194	-.12476
16602	-.41094	-1.68843	.26255	-.35570
16652	-.45118	-.86541	.26255	-.39463
18702	-.27008	-.53620	-.34194	-.28431
19007	1.46045	.45143	.26255	1.64551
19013	1.15861	.28682	.26255	1.60211
19090	1.37996	.45143	.26255	1.68273
19401	1.17873	.45143	.26255	1.82421
22701	.41408	.78064	.26255	-.39767
24354	-.99449	-1.35922	-.34194	-.48064
27101	-.49143	1.43905	-.94644	-.20344
28307	-.73290	.61603	.26255	-.33707
28712	-.55180	.28682	-.94644	-.32125
32347	-1.19572	-.86541	.26255	-.57445
32570	-.45118	-.86541	.26255	-.39147
32601	-.51155	-1.03001	-.34194	-.42529
32812	.21286	-.37159	-.34194	-.18827
36083	-1.17559	-.37159	1.47154	-.43817
37662	-.87376	-1.85303	-.34194	-.35896
40356	-.75302	-.04239	-.34194	-.34700
40505	-.63229	.45143	-.34194	-.34319
43326	-.02861	.45143	.26255	-.39549
45431	.17261	.12222	-.34194	-.00311
55107	1.01776	1.76826	.26255	.44516
61614	.65555	.12222	-.34194	-.37420
62837	-.65241	-.86541	-1.55093	-.53112
74074	-.89388	-.20699	-.94644	-.51668
75149	.83665	1.43905	-.34194	.54225
77701	.17261	-.70080	-.34194	-.39929
83440	-.27008	.78064	2.68053	-.53689
84062	.63543	.61603	2.07604	-.11596
88001	-.69265	-.70080	2.68053	-.47324
91105	1.94339	.28682	.26255	4.09154
92410	1.46045	-.53620	-.34194	.28016
94965	1.96351	.12222	-.94644	.84670
98199	1.23910	-.04239	-1.55093	.45987
99207	.07200	-.37159	-.34194	-.41006

ZIPCODE	ZGOVERN	ZANPSAA	ZSEASONA	ZEILLRAT
2905	.05908	.95345	-1.13453	-.60895
11234	.05908	4.86609	-1.48292	.63225
14850	-.31929	-.15056	-.43774	1.15571
15401	-.69766	.22922	.25904	-.19525
16602	-.31929	-.22122	.60744	.54392
16652	.43746	-.29188	.14291	-.78031
18702	.05908	-.01808	-.78614	-.51213
19007	.05908	2.82881	-1.13453	.22428
19013	-.31929	2.92302	-1.01840	.20529
19090	.05908	2.97896	-1.01840	-.17843
19401	-.31929	3.25275	-1.01840	-.24030
22701	.81583	-.61867	.60744	.33965
24354	-.69766	-.53329	1.07196	-.35516
27101	-1.07603	-.46558	-.78614	-.63392
28307	.43746	-.48618	-.32161	.22428
28712	-.69766	-.41847	-.55387	-.20171
32347	1.19420	-.69816	.60744	5.01003
32570	2.32932	-.47735	-.55387	.95998
32601	1.19420	-.55979	.14291	-.01861
32812	.05908	-.38020	.14291	.26536
36083	1.19420	-.38609	-.67001	1.34155
37662	-.69766	-.40964	.83970	.43680
40356	.05908	-.35370	.14291	.22428
40505	.43746	-.33604	.14291	.20439
43326	-.69766	-.54212	-.32161	-1.72082
45431	.05908	-.06813	-.78614	.32317
55107	-.31929	1.44511	-.90227	.10891
61614	-1.07603	-.14762	.14291	-.12867
62837	-.69766	-.64516	1.18809	.22428
74074	.05908	-.59511	.95583	.22428
75149	-.31929	.64139	-.55387	-.63907
77701	-.69766	-.48618	.37517	-.25134
83440	.05908	-.58628	1.53648	.36849
84062	.05908	.14384	-.43774	.15740
88001	2.32932	-.51857	.25904	-.12621
91105	-.69766	1.89849	-1.25066	.25065
92410	.05908	-.14468	-.55387	.12847
94965	.43746	1.03883	-1.01840	-3.18143
98199	.05908	.38820	-.78614	-.27994
99207	.05908	-.30954	.02678	-.24811

ZIPCODE	ZPOPCHNG	ZPMILAVA	ZWAGES	ZMANUFAC
2905	-.81067	.43007	-.22739	.88745
11234	-1.51242	5.36152	-.06820	-.04903
14850	-.88085	-.31087	.33323	.30215
15401	-.74050	-.10885	-.73956	.30215
16602	-.74050	-.33817	-.55961	.30215
16652	-.45980	-.37820	-.80877	.77039
18702	-.60015	-.27015	-.56653	1.12157
19007	-1.02120	1.76267	.31939	.18509
19013	-1.09137	1.70845	.96999	.30215
19090	-1.09137	1.79334	1.10841	.30215
19401	-1.09137	1.94030	1.10841	.53627
22701	1.50509	-.39711	-.69803	-.63433
24354	.17177	-.48239	-1.25173	.18509
27101	.10160	-.19852	.99075	1.94099
28307	.45247	-.32841	-1.13407	.53627
28712	.24195	-.30907	1.53060	1.12157
32347	.31212	-.58012	.30555	-.40021
32570	.38230	-.38590	-.87798	-.98551
32601	2.83841	-.42056	-.80877	-1.45375
32812	2.27701	-.16208	.40244	-1.33669
36083	.59282	-.41918	-1.63932	.30215
37662	.24195	-.35370	1.01843	.06803
40356	.45247	-.34331	-.85030	.06803
40505	.45247	-.33950	.28478	-.16609
43326	-.45980	-.39381	-.04744	1.35569
45431	-.95102	.01164	.42321	.77039
55107	-.45980	.41752	1.50292	.06803
61614	-.31945	-.37058	.80388	.65333
62837	-.38963	-.53625	-.53884	-.40021
74074	1.08404	-.50687	-.99564	-.75139
75149	1.08404	.46011	1.92511	.06803
77701	-.31945	-.39656	1.46139	.06803
83440	1.22439	-.53831	-.69111	-1.21963
84062	1.85596	-.16380	-.08896	-.75139
88001	2.41736	-.46573	-1.44552	-1.45375
91105	-.24928	3.95266	1.73132	.18509
92410	1.43491	.29039	.14636	-.63433
94965	-.74050	.69836	1.01843	-.86845
98199	.17177	.38389	1.34373	-.16609
99207	.80334	-.41904	-.20662	-.98551

ZIPCODE	ZUNEMPB	ZUNEMPW	ZMILINS	ZACTREC
2905	-.44231	-.16348	.88285	.50552
11234	.60291	-.43721	3.68442	4.91000
14850	-.78047	.52086	-.45956	-.32703
15401	-.13489	1.06832	-.57630	-.03161
16602	-1.05714	2.23169	-.57630	-.46131
16652	-1.11863	1.37627	-.57630	-.48817
18702	-.93418	1.03411	-.34283	-.11218
19007	.69514	-.33456	2.10854	1.76778
19013	.81811	-.30034	1.75834	1.74092
19090	.47995	-.36878	2.05017	1.74092
19401	.44921	-.40299	1.81671	2.38548
22701	1.43294	-.98467	-.28447	-.43445
24354	.57217	.82881	-.57630	-.59559
27101	.91033	-.84781	-.45956	-.46131
28307	1.27923	-.84781	-.40120	-.30017
28712	-.10415	-.53986	-.57630	-.32703
32347	.91033	-1.32684	-.57630	-.67616
32570	1.15627	-.60829	.18246	-.38074
32601	1.67888	-.67672	-.57630	-.46131
32812	.11105	-.88202	-.16773	-.30017
36083	1.34072	-.98467	-.45956	-.48817
37662	-.07341	.58929	-.57630	-.46131
40356	1.00256	.55507	-.45956	-.32703
40505	1.64814	.21291	-.45956	-.32703
43326	-.90343	.24712	-.45956	-.32703
45431	.29550	.58929	-.10937	.29067
55107	-.50379	-.47143	-.45956	.66666
61614	-.99566	.04182	-.45956	-.40760
62837	-1.36456	.96567	-.57630	-.51502
74074	-.41156	-1.25841	-.57630	-.62245
75149	-.31934	-1.39527	-.05100	.47867
77701	.17253	-.74516	-.16773	-.48817
83440	-1.42604	-.26613	-.57630	-.56874
84062	-.19637	-.19769	-.45956	-.30017
88001	1.34072	-.30034	-.45956	-.51502
91105	1.00256	-.33456	1.05795	3.75516
92410	.35698	-.06083	-.10937	.39810
94965	.94107	-.33456	2.57547	.82780
98199	.04956	.07604	1.58324	.10268
99207	-.50379	.99989	-.40120	-.40760

APPENDIX G
ACCESSION CLUSTER STATISTICS

Cluster 1 ARMY NONPRIOR SERVICE MALE & FEMALE ALL

MEAN 852.699 STD DEV 461.500
MINIMUM 133 MAXIMUM 1903
VALID OBSERVATIONS - 146 MISSING - 0

Cluster 2 ARMY NONPRIOR SERVICE MALE & FEMALE ALL

MEAN 92.562 STD DEV 82.579
MINIMUM 6 MAXIMUM 361
VALID OBSERVATIONS - 48 MISSING - 0

Cluster 3 ARMY NONPRIOR SERVICE MALE & FEMALE ALL

MEAN 122.016 STD DEV 122.542
MINIMUM 0 MAXIMUM 730
VALID OBSERVATIONS - 742 MISSING - 0

APPENDIX H
ACCESSION COUNT ANALYSIS OF VARIANCE CLUSTER VARIABLES

ANALYSIS OF VARIANCE.

VARIABLE	CLUSTER MS	DF	ERROR MS	DF	F	PROB
ZUNEMPB	56.8467	2	.8666	933.0	65.5975	.000
ZAVGFAMS	191.6622	2	.6075	933.0	315.5144	.000
ZMANUFAC	14.1127	2	.9742	933.0	14.4859	.000
ZWAGES	85.3990	2	.8191	933.0	104.2620	.000
ZPOPCHNG	88.2305	2	.8200	933.0	107.6024	.000
ZMILINS	275.5586	2	.3914	933.0	704.0553	.000
ZPMILAVA	255.9610	2	.4490	933.0	570.0828	.000

APPENDIX I
ACCESSION DISTANCE COEFFICIENTS

AGGLOMERATION SCHEDULE USING AVERAGE LINKAGE (BETWEEN GROUPS)

	CLUSTERS COMBINED			STAGE CLUSTER 1ST APPEARS		NEXT
STAGE	CLUST 1	CLUST 2	COEFFICIENT	CLUST 1	CLUST 2	STAGE
1	10	11	.149903	0	0	2
2	9	10	.283479	0	1	6
3	5	6	.493985	0	0	12
4	13	23	.526509	0	0	8
5	22	32	.710301	0	0	14
6	8	9	.798603	0	2	28
7	7	25	.861992	0	0	12
8	13	15	1.200522	4	0	24
9	3	28	1.221478	0	0	17
10	17	24	1.302960	0	0	21
11	30	40	1.310357	0	0	23
12	5	7	1.462651	3	7	17
13	12	18	1.633371	0	0	21
14	22	27	1.660768	5	0	18
15	20	37	1.702411	0	0	23
16	4	26	2.268536	0	0	20
17	3	5	2.280144	9	12	22
18	22	31	2.648871	14	0	29
19	14	16	2.752981	0	0	29
20	1	4	3.005300	0	16	22
21	12	17	3.011827	13	10	24
22	1	3	3.151413	20	17	30
23	20	30	3.341523	15	11	31
24	12	13	3.391588	21	8	27
25	33	34	3.589299	0	0	35
26	38	39	4.019897	0	0	32
27	12	21	5.439097	24	0	31
28	8	36	5.448341	6	0	32
29	14	22	6.152546	19	18	30
30	1	14	6.985703	22	29	33
31	12	20	7.244295	27	23	34
32	8	38	8.211899	28	26	37
33	1	29	8.884380	30	0	36
34	12	19	9.122895	31	0	36
35	33	35	9.210579	25	0	39
36	1	12	11.836977	33	34	38

37	2	8	15.776648	0	32	38
38	1	2	19.065002	36	37	39
39						

LIST OF REFERENCES

1. Brinkerhoff, J. and Grissmer, D., *The Reserve Forces in an All Volunteer Environment*, Rand, Santa Monica, CA, January 1984.

2. Defense Manpower Data Center, *Official Guard and Reserve Manpower Strengths and Statistics*, RCS: DD M(M) 1147 1148, Washington DC, September 1985.

3. Congressional Budget Office, *Improving the Army Reserve*, Congress of the United States, Washington DC, November 1985.

4. Gentlemen, J., *Computer Science and Statistics: 12th Annual Symposium on the Interface*, University of Waterloo, Waterloo, Ontario, Canada, May 10 & 11, 1979.

5. Dertouzos, J., *Recruiter Incentives and Enlistment Supply*, Rand R-3065-MIL, Santa Monica, CA, May 1985.

6. Goldberg, L., *Enlisted Supply: Past, Present and Future*, CNS 1168-vol 1, Centre for Naval Analysis, VA, September 1985.

7. Daula, T. and Smith, D., *Recruiting Goals, Enlistment Supply and Enlistments in the U.S. Army*, U.S. Army Manpower Economics Conference, Virginia, December 1984.

8. Golberg, L., and Greenston, J., *Analysis of Enlistment Supply in FY 1976-1983*, Economic Research Laboratory, Reston, VA, 1984.

9. Kelly, R., *The Supply of Volunteers to the Selected Reserve*, Department of Social Sciences, USMA, May 1979.

10. Rostker, B., *Air Reserve Personnel Study: Volume III. Total Force Planning Personnel Costs and the Supply of New Reservists*, Rand R-1430-PR, Santa Monica, CA, October 1974.

11. McNaught, W., *The Supply of Enlistees to the Selected Reserve Forces*, Rand N-1562-MRL, Santa Monica, CA, July 1981.

12. Borack, J., Mehay, S. and Thomas, G., *Estimating Local Area Manpower Supply for the Reserves*, NPS 54-85-006, Department Administrative Sciences, Naval Postgraduate School, Monterey,CA, September 1985.

13. Grissmer, D., and Kirby, S., *Attrition and Retention in the Army Reserve and National Guard: An Empirical Analysis*, Rand Corporation, P-7077, Santa Monica, CA, March 1985.

14. Aldenderfer, M., and Blashfield, R., *Cluster Analysis*, Sage Publications, Beverly Hills, CA, 1984.

15. Solomon, H., *Numerical Taxonomy*, Technical Report, ONR NR-042-267, Stanford University, December 1970.

16. Ball, G., *Classification Analysis*, Technical Report, SRI 5533, Stanford Research Institute, November 1970.

17. Alexander, A., *Income Experience, and the Structure of Internal Labor Markets* Quarterly Journal of Economics, February 1974, 63-85.

18. Milligan, G., *An Examination of the Effect of Six Types of Error Perturbation on Fifteen Clustering Algorithms*, Psychometrika-vol. 45 no. 3, September 1980, 325-342.

19. Hodson, R., *Workers' Earnings and Corporate Economic Structure*, Academic Press, NY, 1983.

20. Kotz, S., and Cambell, R., *Encyclopedia of Statistical Sciences*, Volume 2, Classification to Eye Estimate, Wiley & Sons, NY, 1982.

21. Kuiper, F. and L. Fisher, *A Montecarlo comparison of Six Clustering Procedures*, Biometrics, 1975.

22. Skinner, H., *Dimensions and Clusters: A Hybrid Approach to Classification*, Applied Psychological Measurement, 1979.

23. Osterman, P., *An Empirical Study of Labor Market Segmentation*, Industrial and Labor Relations Review, July 1975, 508-23.

24. Pindyck and Rubinfeld, *Econometric Models and Economic Forecasts*, McGraw-Hill, 1981.

25. Department of Defense Instruction 7730.54, *Reserve Components Common Personnel Data System*, Washington DC, October 1981.

26. Department of the Army Regulation 140-1, *Army Reserve: Mission Organization and Training*, Washington DC, 1 April 1983.

27. Shisko, R. and Rostker, B., *The Economics of Multiple Job Holding*, The American Economic Review, v. LXVI, June 1976, 298-309.

28. Hanssens, D.M., Levien, H.A., *An Econometric Study of Recruitment Marketing in the U.S. Navy*, Management Science, Vol. 29, No. 10, October 1983, 167-182.

29. Bartholomew, D.J., and Forbes, A.F., *Statistical Techniques for Manpower Planning*, John Wiley and Sons, 1979.

30. Anderberg, M., *Cluster Analysis for Applications*, Academic Press, NY, 1973.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, VA 22304-6145	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, CA 93943-5002	2
3.	Department of the Navy Chief of Naval Operation (OP-114) ATTN: LCDR Cromwell Washington, DC 20350-2000	1
4.	Professor George W. Thomas Code 54TE Naval Postgraduate School Monterey, CA 93943	4
5.	Professor Stephen L. Mehay Code 54MP Naval Postgraduate School Monterey, CA 93943	2
6.	Alvin R. Jones 3319 Ripple Road Baltimore, MD 21207	3

END

10-87

DTIC